POLYPEPTIDES THAT BIND HIV gp120 AND RELATED NUCLEIC ACIDS, ANTIBODIES, COMPOSITIONS, AND METHODS OF USE

TECHNICAL FIELD OF THE INVENTION

The present invention relates to polypeptides with homology to regions of domains of the human chemokine receptors CCR5, CXCR4, and STRL33, as well as domains of CD4 that bind with human immunodeficiency virus (HIV), in particular HIV-1 glycoprotein 120 (gp120) envelope protein. The present invention also relates to nucleic acids encoding such polypeptides, antibodies, compositions comprising such polypeptides, nucleic acids or antibodies, and methods of using the same.

BACKGROUND OF THE INVENTION

There are seven transmembrane chemokine receptors that act as cofactors for HIV infection. The cofactors enable entry of HIV-1 into $CD4^{+}$ T cells and macrophages (Premack et al., Nature Medicine 2: 1174-78 (1996); and Zhang et al., Nature 383: 768 (1996)).

The presence of chemokines has an inhibitory effect on HIV-1 attachment to, and infection of, susceptible cells. Additionally, some mutations in chemokine receptors have been shown to result in resistance to HIV-1 infection. For example, a 32-nucleotide deletion within the CCR5 gene has been described in subjects who remained uninfected despite repeated exposures to HIV-1 (Huang et al., Nature Medicine 2: 1240-43 (1996)).

Evidence also exists for the physical association of a ternary complex between chemokine receptors, CD4, and HIV-1 gp120 envelope glycoprotein on cell membranes (Lapham et al., Science 274: 602-05 (1996)). Receptor signaling and cell activation are probably not required for the anti-HIV-1 effect of chemokines since a RANTES analog lacking the first eight amino-terminal amino acids, RANTES

(9-68), lacked chemotactic and leukocyte-activating properties, but bound to multiple chemokine receptors and inhibited infection by macrophage-tropic HIV-1 (Arenzana-Seladedos et al., Nature 383: 400 (1996)). Cumulatively, the above described results suggest that the interaction between gp120, CD4, and at least one chemokine receptor is obligatory for HIV-1 infection. Accordingly, reagents that interfere with the binding of gp120 to chemokine receptors and to CD4 are used in the biological and medical arts. However, there presently exists a need for additional reagents that can compete with one or more proteins of the gp120-CD4-chemokine receptor complex to assist in basic biological or viral research, and to assist in medical intervention in the HIV-1 pandemic. It is an object of the present invention to provide such reagents. This and other objects and advantages, including additional inventive features, will be apparent from the description provided herein.

BRIEF SUMMARY OF THE INVENTION

The present invention provides a polypeptide that binds with HIV gp120 under physiological conditions. Multiple embodiments of the present inventive polypeptide are provided, and each embodiment possesses a degree of homology to at least one of the human CCR5, CXCR4 and STRL33 chemokine receptors, and the human CD4 cell-surface protein.

In a first embodiment, the present invention provides a polypeptide comprising the amino acid sequence YDIXYYXXE (SEQ ID NO: 1), wherein X is any synthetic or naturally occurring amino acid residue, and the polypeptide comprises less than about 100 contiguous amino acids that are identical to, or, in the alternative, substantially identical to, the amino acid sequence of the human CCR5 chemokine receptor. A preferred polypeptide of this first

embodiment comprises the amino acid sequence YDIN*YYT*S*E (SEQ ID NO: 3). A more preferred polypeptide of this first embodiment comprises the amino acid sequence YDINYYTSE (SEQ ID NO: 3), wherein each letter is the standard one-letter abbreviation for an amino acid residue (i.e., for example, N denotes asparaginyl, T denotyes threoninyl, and S denotes serinyl). The polypeptide of the first embodiment can comprise the amino acid sequence M*D*YQ*V*S*SP*IYDIN*YYT*S*E (SEQ ID NO: 5). Preferably, the polypeptide comprises the amino acid sequence MDYQVSSPIYDINYYTSE (SEQ ID NO: 5).

In a second embodiment, the present invention provides a polypeptide comprising the amino acid sequence XEXIXIYXXXNYXXX (SEQ ID NO: 6), wherein X is any synthetic or naturally occurring amino acid and wherein said polypeptide comprises less than about 100 contiguous amino acid that are identical to or substantially identical to the amino acid sequence of the human CXCR4 chemokine receptor. The polypeptide can consist essentially of, or consist of, the sequence EXIXIYXXXNY (SEQ ID NO: 7). Preferably, the polypeptide comprises the sequence M*EG*IS*IYT*S*D*NYT*E*E*. Preferably, M*EG*IS*IYT*S*D*NYT*E*E* is M*EGISIYTSDNYT*E*E*.

In a third embodiment, the present invention provides a polypeptide comprising the amino acid sequence EHQAFLQFS (SEQ ID NO: 10), wherein said polypeptide comprises less than about 100 contiguous amino acids that are identical to or substantially identical to the amino acid sequence of the human STRL33 chemokine receptor. The polypeptide can consist essentially of, or consist of, the sequence EHQAFLQFS (SEQ ID NO: 10).

In a fourth embodiment, the present invention provides a polypeptide comprising at least a portion of an amino acid sequence selected from the group consisting of LPPLYSLVFIFGFVGNML (SEQ ID NO: 11), QWDFGNTMCQLLTGLYFIGFFS

(SEQ ID NO: 12), SQYQFWKNFQTLKIVILG (SEQ ID NO: 13), APYNIVLLLNTFQEFFGLNNCS (SEQ ID NO: 14), and YAFVGEKFRNYLLVFFQK (SEQ ID NO: 15), wherein said polypeptide comprises less than about 100 contiguous amino acids that are identical to or substantially identical to the amino acid sequence of the human CCR5 chemokine receptor.

In a fifth embodiment, the present invention provides a polypeptide comprising at least a portion of an amino acid sequence selected from the group consisting of LLLTIPDFIFANVSEADD (SEQ ID NO: 16), VVFQFQHIMVGLILPGIV (SEQ ID NO: 17), and IDSFILLEIIKQGCEFEN (SEQ ID NO: 18), wherein said polypeptide comprises less than about 100 contiguous amino acids that are identical to or substantially identical to the amino acid sequence of the human CXCR4 chemokine receptor.

In a sixth embodiment, the present invention provides a polypeptide comprising at least a portion of an amino acid sequence selected from the group consisting of LVISIFYHKLQSLTDVFL (SEQ ID NO: 19), PFWAYAGIHEWVFGQVMC (SEQ ID NO: 20), EAISTVVLATQMTLGFFL (SEQ ID NO: 21), LTMIVCYSVIIKTLLHAG (SEQ ID NO: 22), MAVFLLTQMPFNLMKFIRSTHW (SEQ ID NO: 23), HWEYYAMTSFHYTIMVTE (SEQ ID NO: 24), ACLNPVLYAFVSLKFRKN (SEQ ID NO: 25) and SKTFSASHNVEATSMFQL (SEQ ID NO: 26), wherein said polypeptide comprises less than about 100 contiguous amino acids that are identical to or substantially identical to the amino acid sequence of the human STRL33 chemokine receptor.

In a seventh embodiment, the present invention provides a polypeptide comprising at least a portion of an amino acid sequence selected from the group consisting of DTYICEVED (SEQ ID NO: 27), EEVQLLVFGLTANSD (SEQ ID NO: 28), THLLQGQSLTLTLES (SEQ ID NO: 29), and GEQVEFSFPLAFTVE (SEQ ID NO: 30), wherein said polypeptide comprises less than about 100 contiguous amino acids that are identical to or

substantially identical to the amino acid sequence of the human CD4 cell-surface protein.

In the fourth to seventh embodiments, any selected portion of the polypeptide can comprise from 1 to about 6 conservative amino acid substitutions. In an alternative, the polypeptide can be partially defined by an absence of a polypeptide sequence, outside the region of the portion selected from the foregoing sequences, that has five, or ten, contiguous amino acid residues that have a sequence that consists of an amino acid sequence that is identical to or substantially identical to the protein to which the polypeptide has homology (i.e., CCR5, CXCR4, STRL33, or CD4). In yet another alternative, the polypeptide can lack a sequence of five or ten contiguous amino acids which are identical to or substantially identical to the sequence of the protein with which the sequence has homology except that one or more conservatively or neutrally substituted amino acids replace part of the sequence of the protein to which the polypeptide has homology. Additionally, any embodiment of the present inventive polypeptide can also comprise a pharmaceutically acceptable substituent.

Any embodiment of the present inventive polypeptide can be incorporated into a composition, which further comprises a carrier. Any suitable embodiment of the present inventive polypeptide can be encoded by a nucleic acid that can be expressed in a cell. In this regard, the present invention further provides a vector comprising such a nucleic acid. The nucleic acids and vectors also can be incorporated into a composition comprising a carrier.

Additionally, the present invention provides a method of making an antibody to a polypeptide of the present invention. The present invention also provides a method of prophylactically or therapeutically treating an HIV infection in a mammal.

Additionally, the present invention provides an anti-idiotypic antibody comprising an internal image of a portion of gp120, as well as a method of selecting such an antibody.

The present invention also provides a method of making an antibody to a portion of the gp120 protein that binds with a portion of CCR5, CXCR4, STRL33, or CD4, as well as the immunizing compound used to make the antibody, and the antibody itself. In another embodiment of the present invention, a method of removing HIV-1 from a bodily fluid is provided.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 depicts a listing of synthetic amino acids available (from Bachem, King of Prussia, PA) for incorporation into polypeptides of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a polypeptide that binds with gp120 of HIV, in particular HIV-1, more particularly HIV-1_{LAI}, under physiological conditions. The polypeptide has a number of uses including, but not limited to, the use of the polypeptide to elucidate the mechanism by which HIV, such as HIV-1, attaches to and/or infects a particular cell, to induce an immune response in a mammal, in particular a human, to HIV, in particular HIV-1, and to inhibit the replication of HIV, in particular HIV-1, in an infected mammal, in particular a human.

Multiple embodiments of the present inventive polypeptide are provided. Each embodiment of the polypeptide has a degree of homology to at least one of the human CCR5, CXCR4 and STRL33 chemokine receptors, or the human CD4 cell-surface protein. In each embodiment provided herein, a letter indicates the standard amino acid designated by that letter, and a letter followed directly

by an asterisk (*) preferably represents the amino acid represented by the letter (e.g., N represents asparaginyl and T represents threoninyl), or a synthetic or naturally occurring conservative or neutral substitution therefor. Additionally, in accordance with convention, all amino acid sequences provided herein are given either from left to right, or top to bottom, such that the first amino acid is amino-terminal and the last is carboxyl-terminal. The synthesis of polypeptides, either synthetically (i.e., chemically) or biologically, is within the skill in the art.

It is within the skill of the ordinary artisan to select synthetic and naturally occurring amino acids that make conservative or neutral substitutions for any particular naturally occurring amino acids. The skilled artisan desirably will consider the context in which any particular amino acid substitution is made, in addition to considering the hydrophobicity or polarity of the side-chain, the general size of the side chain, and the pK value of side-chains with acidic or basic character under physiological conditions. For example, lysine, arginine, and histidine are often suitably substituted for each other, and more often arginine and lysine. As is known in the art, this is because all three amino acids have basic side chains, whereas the pK value for the side-chains of lysine and arginine are much closer to each other (about 10 and 12) than to histidine (about 6). Similarly, glycine, alanine, valine, leucine, and isoleucine are often suitably substituted for each other, with the proviso that glycine is frequently not suitably substituted for the other members of the group. This is because each of these amino acids are relatively hydrophobic when incorporated into a polypeptide, but glycine's lack of an α-carbon allows the phi and psi angles of rotation (around the α-carbon) so

much conformational freedom that glycinyl residues can trigger changes in conformation or secondary structure that do not often occur when the other amino acids are substituted for each other. Other groups of amino acids frequently suitably substituted for each other include, but are not limited to, the group consisting of glutamic and aspartic acids; the group consisting of phenylalanine, tyrosine and tryptophan; and the group consisting of serine, threonine and, optionally, tyrosine. Additionally, the skilled artisan can readily group synthetic amino acids with naturally occurring amino acids.

In the context of the present invention, a polypeptide is "substantially identical" to another polypeptide if it comprises at least about 80% identical amino acids. Desirably, at least about 50% of the non-identical amino acids are conservative or neutral substitutions. Also, desirably, the polypeptides differ in length (i.e., due to deletion mutations) by no more than about 10%.

In a first embodiment, the present invention provides a polypeptide comprising the amino acid sequence YDIXYYXXE (SEQ ID NO: 1), wherein X is any synthetic or naturally occurring amino acid residue, and the polypeptide comprises less than about 100 contiguous amino acids, preferably less than about 50 amino acids, more preferably less than about 25 amino acids, and yet more preferably less than about 13 amino acids that are identical to, or, in the alternative, substantially identical to, the amino acid sequence of the human CCR5 chemokine receptor.

Preferably, the polypeptide of the first embodiment comprises YDIXYYXXE (SEQ ID NO: 1), wherein the amino moiety of the amino-terminal tyrosinyl residue is not bound to another amino acid residue via a peptidic bond, and the carboxyl moiety of the glutamyl residue is not bound to another amino acid residue via a peptidic bond. However, the polypeptide can consist essentially of YDIXYYXXE (SEQ

ID NO: 1) and, optionally, can be modified by one or more pharmaceutically acceptable substituents, such as, for example, t-boc or a saccharide.

More particularly, the polypeptide comprises the amino acid sequence YDIN*YYT*S*E (SEQ ID NO: 3). Preferably, N* is asparaginyl, T* is threoninyl, and S* is serinyl.

The polypeptide of the first embodiment can comprise a dodecapeptide selected from the amino acid sequence M*D*YQ*V*S*SP*IYDIN*YYT*S*E (SEQ ID NO: 5). More preferably, the polypeptide of the first embodiment comprises the amino acid sequence MDYQVSSPIYDINYYTSE (SEQ ID NO: 5).

In a second embodiment, the present invention provides a polypeptide comprising the amino acid sequence XEXIXIYXXXNYXXX (SEQ ID NO: 6), wherein X is any synthetic or naturally occurring amino acid, and the polypeptide comprises less than about 100 contiguous amino acids, preferably less than about 50 amino acids, and more preferably less than about 25 amino acids, that are identical to or substantially identical to the amino acid sequence of the human CXCR4 chemokine receptor. Optionally, the polypeptide consists essentially of, or consists of, the sequence EXIXIYXXXNY (SEQ ID NO: 7).

In a preferred polypeptide of this second embodiment, the polypeptide comprises the amino acid sequence M*EG*IS*IYT*S*D*NYT*E*E*. Preferably, M*EG*IS*IYT*S*D*NYT*E*E* is M*EGISIYTSDNYT*E*E*.

In a third embodiment, the present invention provides a polypeptide comprising the amino acid sequence EHQAFLQFS, wherein the polypeptide comprises less than about 100 contiguous amino acid residues, preferably less than about 50 contiguous amino acid residues, more preferably less than about 25 contiguous amino acid residues, that are identical to or substantially identical to the amino acid sequence of the human STRL33 chemokine receptor. The

polypeptide can consist essentially of, or consist of, the sequence EHQAFLQFS.

The first three embodiments of the present invention provide, among other things, polypeptides having substantial identity or identity to the amino-terminal regions of the chemokine receptors CCR5, CXCR4, and STRL33. These first three embodiments form a first group of embodiments of the present invention. The present invention also provides, in a second group of embodiments, polypeptides having substantial identity or identity to an internal region of the human chemokine receptors CCR5, CXCR4, and STRL33, as well as to the leukocyte cell-surface protein CD4.

This second group of embodiments provides a polypeptide that binds with HIV gp120 under physiological conditions and comprises at least a portion of or all of an amino acid sequence selected from the group consisting of LPPLYSLVFIFGFVGNML (SEQ ID NO: 11), QWDFGNTMCQLLTGLYFIGFFS (SEQ ID NO: 12), SQYQFWKNFQTLKIVILG (SEQ ID NO: 13), APYNIVLLLNTFQEFFGLNNCS (SEQ ID NO: 14), and YAFVGEKFRNYLLVFFQK (SEQ ID NO: 15), wherein the polypeptide comprises less than about 100 amino acids that are identical to or substantially identical to the amino acid sequence of the human CCR5 chemokine receptor; or selected from the group consisting of LLLTIPDFIFANVSEADD (SEQ ID NO: 16) (165-182), VVFQFQHIMVGLILPGIV (SEQ ID NO: 17) (197-214), and IDSFILLEIIKQGCEFEN (SEQ ID NO: 18) (261-278), wherein the polypeptide comprises less than about 100 amino acids that are identical to or substantially identical to the amino acid sequence of the human CXCR4 chemokine receptor; or

selected from the group consisting of LVISIFYHKLQSLTDVFL (SEQ ID NO: 19) (53-70), PFWAYAGIHEWVFGQVMC (SEQ ID NO: 20) (85-102), EAISTVVLATQMTLGFFL (SEQ ID NO: 21) (185-202),

LTMIVCYSVIIKTLLHAG (SEQ ID NO: 22) (205-222), MAVFLLTQMPFNLMKFIRSTHW (SEQ ID NO: 23) (237-258), HWEYYAMTSFHYTIMVTE (SEQ ID NO: 24) (257-274), ACLNPVLYAFVSLKFRKN (SEQ ID NO: 25) (281-298) and SKTFSASHNVEATSMFQL (SEQ ID NO: 26) (325-342), wherein the polypeptide comprises less than about 100 amino acids that are identical to a substantially identical to the amino acid sequence of the human STRL33 chemokine receptor; or

selected from the group consisting of DTYICEVED (SEQ ID NO: 27), EEVQLLVFGLTANSD (SEQ ID NO: 28), THLLQGQSLTLTLES (SEQ ID NO: 29), and GEQVEFSFPLAFTVE (SEQ ID NO: 30), wherein the polypeptide binds with HIV gp120 under physiological conditions and comprises less than about 100 amino acids that are identical to or substantially identical to the amino acid sequence of the human CD4 cell-surface protein. Optionally, the recited amino acid sequences can comprise 1 to about 6 conservative or neutral amino acid substitutions.

The polypeptides of this second group of embodiments preferably comprise less than about 50 amino acid residues, and more preferably less than about 25 amino acid residues, and yet more preferably no additional amino acid residues, that are identical to a protein that naturally has the recited amino acid sequence. The polypeptide can be alternatively characterized by an absence of a region, outside the above-recited amino acid sequences, that has about five, or about ten, contiguous amino acid residues that have a sequence that consists of an amino identical and conservatively substituted residues as an amino acid sequence of the protein to which the polypeptide of the compound has homology.

Any embodiment of the present inventive polypeptide can also comprise a pharmaceutically acceptable substituent, attachment of which is within the skill in the art. The pharmaceutically acceptability of substituents

are understood by those skilled in the art. For example, a pharmaceutically acceptable substituent can be a biopolymer, such as a polypeptide, an RNA, a DNA, or a polysaccharide. Suitable polypeptides comprise fusion proteins, an antibody or fragment thereof, a cell adhesion molecule or a fragment thereof, or a peptide hormone. Suitable polysaccharides comprise polyglucose moieties, such as starch and their derivatives, such as heparin. The pharmaceutically acceptable substituent also can be any suitable lipid or lipid-containing moiety, such as a lipid of a liposome or a vesicle, or even a lipophilic moiety, such as a prostaglandin, a steroid hormone, or a derivative thereof. Additionally, the pharmaceutically acceptable substituent can be a nucleotide or nucleoside, such as nicotine adenine dinucleotide or thymine, an amino acid residue, a saccharide or disaccharide, or the residue of another biomolecule naturally occurring in a cell, such as inositol, a vitamin, such as vitamin C, thiamine, or nicotinic acid. Synthetic organic moieties also can be pharmaceutically acceptable substituents, such as t-butyl carbonyl, an acetyl moiety, quinine, polystyrene and other biologically acceptable polymers. Optionally, a pharmaceutically acceptable substituent can be selected from the group consisting of a C_1-C_{18} alkyl, a C_2-C_{18} alkenyl, a C_2-C_{18} alkynyl, a C_6-C_{18} aryl, a C_7-C_{18} alkaryl, a C_7-C_{18} aralkyl, and a C_3-C_{18} cycloalkyl, wherein any of the foregoing moieties that are cyclic comprise from 0 to 2 atoms per carbocyclic ring, which can be the same or different, and are selected from the group consisting of nitrogen, oxygen, and sulfur.

Any of the substituents from this group can be substituted by one to six substituent moieties, which can be the same or different, selected from the group consisting of an amino moiety, a carbamate moiety, a carbonate moiety, hydroxyl, a phosphamate moiety, a

phosphate moiety, a phosphonate moiety, a pyrophosphate moiety, a triphosphate moiety, a sulfamate moiety, a sulfate moiety, a sulfonate moiety, a C_1-C_8 monoalkylamine moiety, a C_1-C_8 dialkylamine moiety, and a C_1-C_8 trialkylamine moiety.

Any embodiment of the present inventive polypeptide can be encoded by a nucleic acid and can be expressed in a cell. The skilled artisan will recognize that the encoded polypeptide as well as any pharmaceutically acceptable substituent to be incorporated into the polypeptide, e.g., a formyl or acetyl substituent on an amino-terminal methionine or a saccharide, will preferably be produced by a cell that can express the polypeptide of the present invention. Accordingly, the amino acids incorporated into the polypeptide encoded by the nucleic acid are preferably naturally occurring.

A nucleic acid as described above can be cloned into any suitable vector and can be used to transduce, transform, or transfect any suitable host. The selection of vectors and methods to construct them are commonly known to persons of ordinary skill in the art and are described in general technical references (see, in general, "Recombinant DNA Part D," Methods in Enzymology, Vol. 153, Wu and Grossman, eds., Academic Press (1987)). Desirably, the vector comprises regulatory sequences, such as transcription and translation initiation and termination codons, which are specific to the type of host (e.g., bacterium, fungus, plant, or animal) into which the vector is to be inserted, as appropriate and taking into consideration whether the vector is DNA or RNA. Preferably, the vector comprises regulatory sequences that are specific to the genus of the host. Most preferably, the vector comprises regulatory sequences that are specific to the species of the host and is optimized for the expression of an above-described polypeptide.

Constructs of vectors, which are circular or linear, can be prepared to contain an entire nucleic acid sequence as described above or a portion thereof ligated to a replication system that is functional in a prokaryotic or eukaryotic host cell. Replication systems can be derived from ColE1, 2 mμ plasmid, λ, SV40, bovine papilloma virus, and the like.

Suitable vectors include those designed for propagation and expansion, or for expression, or both. A preferred cloning vector is selected from the group consisting of the pUC series, the pBluescript series (Stratagene, LaJolla, CA), the pET series (Novagen, Madison, WI), the pGEX series (Pharmacia Biotech, Uppsala, Sweden), and the pEX series (Clonetech, Palo Alto, CA). Examples of animal expression vectors include pEUK-C1, pMAM and pMAMneo (Clonetech, Palo Alto, CA).

An expression vector can comprise a native or nonnative promoter operably linked to a nucleic acid molecule encoding an above-described polypeptide. The selection of promoters, e.g., strong, weak, inducible, tissue-specific and developmental-specific, is within the skill in the art. Similarly, the combining of a nucleic acid molecule as described above with a promoter is also within the skill in the art.

The skilled artisan will also recognize that the polypeptide has ability to bind the gp120 protein, which is most often found outside of cells. Accordingly, the present inventive nucleic acid advantageously can comprise a nucleic acid sequence that encodes a signal sequence such that a signal sequence is translated as a fusion protein with the polypeptide of the present inventive polypeptide to form a signal sequence-polypeptide fusion. The signal sequence can cause secretion of the entire polypeptide, including the signal sequence (which is a pharmaceutically acceptable substituent), or can be cleaved from the

polypeptide (i.e., the polypeptide of the compound) prior to, or during, secretion so that at least the present inventive polypeptide is secreted out of a cell in which the nucleic acid is expressed.

Alternatively, the nucleic acid comprises or encodes an antisense nucleic acid molecule or a ribozyme that is specific for a specified amino acid sequence of an above-described polypeptide. A nucleic acid sequence introduced in antisense suppression generally is substantially identical to at least a portion of the endogenous gene or gene to be repressed, but need not be identical. Thus, the vectors can be designed such that the inhibitory effect applies to other proteins within a family of genes exhibiting homology or substantial homology to the target gene. The introduced sequence also need not be full-length relative to either of the primary transcription product or the fully processed mRNA. Generally, higher homology can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments will be equally effective.

Ribozymes also have been reported to have use as a means to inhibit expression of endogenous genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered and is, thus, capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs. The design and use of target RNA-specific ribozymes is described in Haseloff et al., Nature 334: 585-591 (1988).

Further provided by the present invention is a composition comprising an above-described polypeptide or nucleic acid and a carrier therefor. Another composition provided by the present invention is a composition comprising an antibody to an above-described polypeptide or an anti-antibody to an above-described polypeptide.

Any embodiment of the present invention including the present inventive polypeptide, nucleic acid, antibody, and anti-antibody, can be incorporated into a composition comprising a carrier. The carrier can serve any function. For example, the carrier can increase the solubility of the present inventive polypeptide, nucleic acid or antibody in aqueous solutions. Additionally, the carrier can protect the present inventive polypeptide, nucleic acid or antibody from environmental insults, such as dehydration, oxidation, and photolysis. Moreover, the carrier can serve as an adjuvant, or as a timed-release control means in a biological system.

Antibodies can be generated in accordance with methods known in the art. See, for example, Benjamin, In Immunology: a short course, Wiley-Liss, NY, 1996, pp. 436-437; Kuby, In Immunology, 3rd. ed., Freeman, NY, 1997, pp. 455-456; Greenspan et al., FASEB J. 7: 437-443 (1993); and Poskitt, Vaccine 9: 792-796 (1991). Anti-antibodies (i.e., anti-idiotypic antibodies) also can be generated in accordance with methods known in the art (see, for example, Benjamin, In Immunology: a short course, Wiley-Liss, NY, 1996, pp. 436-437; Kuby, In Immunology, 3rd. ed., Freeman, NY, 1997, pp. 455-456; Greenspan et al., FASEB J., 7, 437-443, 1993; Poskitt, Vaccine, 9, 792-796, 1991; and Madiyalakan et al., Hybridonor 14: 199-203 (1995) ("Anti-idiotype induction therapy")). Such antibodies can be obtained and employed either in solution-phase or coupled to a desired solid-phase matrix. Having in hand such antibodies, one skilled in the art will further appreciate

that such antibodies, using well-established procedures (e.g., such as described by Harlow and Lane (1988, supra), are useful in the detection, quantification, or purification of gp120 or HIV, particularly HIV-1, conjugates of each and host cells transformed to produce a gp120 receptor or a derivative thereof. Such antibodies are also useful in a method of prevention or treatment of a viral infection and in a method of inducing an immune response to HIV as provided herein.

In view of the above, an above-described polypeptide can be administered to an animal. The animal generates anti-polypeptide antibodies. Among the anti-polypeptide antibodies generated or induced in the animal are antibodies that have an internal image of gp120. In accordance with well-known methods, polyclonal or monoclonal antibodies can be obtained, isolated and selected. Selection of an anti-polypeptide antibody that has an internal image of gp120 can be based upon competition between the anti-polypeptide antibody and gp120 for binding to an above-described polypeptide, or upon the ability of the anti-polypeptide antibody to bind to a free polypeptide as opposed to a polypeptide bound to gp120. Such an anti-antibody can be administered to an animal to prevent or treat an HIV infection in accordance with methods provided herein.

Although nonhuman anti-idiotypic antibodies, such as an anti-polypeptide antibody that has an internal image of gp120 and, therefore, is anti-idiotypic to gp120, are useful for prophylaxis in humans, their favorable properties might, in certain instances, can be further enhanced and/or their adverse properties further diminished, through "humanization" strategies, such as those recently reviewed by Vaughan, Nature Biotech., 16, 535-539, 1998.

Prior to administration to an animal, such as a mammal, in particular a human, an above-described polypeptide, nucleic acid, antibody or anti-antibody can be formulated into various compositions by combination with appropriate carriers, in particular, pharmaceutically acceptable carriers or diluents, and can be formulated to be appropriate for either human or veterinary applications.

The present invention also provides a method of making an antibody. The method comprises administering an immunogenic amount of an above-described polypeptide or nucleic acid to an animal, such as a mammal, in particular a human. Determining the quantity of a polypeptide or nucleic acid that is immunogenic will depend in part on the degree of similarity to a protein or other molecule of the inoculated animal, the route of administration of the polypeptide or nucleic acid, and the size of the polypeptide administered or encoded by the administered nucleic acid. If necessary, the polypeptide or nucleic acid can be mixed with or ligated to a substance (or an adjuvant) that enhances its immunogenicity. Such calculations and procedures are within the skill of the ordinary artisan. Additionally, the present inventive method preferably can be used to induce an immune response against HIV, particularly HIV-1, in a mammal, particularly a human.

In view of the above, the present invention further provides a method of prophylactically or therapeutically treating an HIV infection in a mammal, particularly a human, in need thereof. The method comprises administering to the mammal an HIV replication-inhibiting effective amount of an above-described polypeptide, nucleic acid, or an anti-antibody to an above-described polypeptide or a nucleic acid encoding such a polypeptide.

The present invention also provides a method of prophylactically or therapeutically treating HIV infection

in a mammal. The method comprises administering to the mammal an effective amount of an above-described polypeptide or nucleic acid. Prior to administration to an animal, such as a mammal, in particular a human, an above-described polypeptide or nucleic acid can be formulated into various compositions by combination with appropriate carriers, in particular, pharmaceutically acceptable carriers or diluents, and can be formulated to be appropriate for either human or veterinary applications.

Thus, a composition for use in the method of the present invention can comprise one or more of the polypeptides, nucleic acids, antibodies or anti-antibodies described herein, preferably in combination with a pharmaceutically acceptable carrier. Pharmaceutically acceptable carriers are well-known to those skilled in the art, as are suitable methods of administration. The choice of carrier will be determined, in part, by whether a polypeptide or a nucleic acid is to be administered, as well as by the particular method used to administer the composition. Optionally, the carrier can be selected to increase the solubility of the composition or mixture, e.g., a liposome or polysaccharide. One skilled in the art will also appreciate that various routes of administering a composition are available, and, although more than one route can be used for administration, a particular route can provide a more immediate and more effective reaction than another route. Accordingly, there are a wide variety of suitable formulations of compositions that can be used in the present inventive methods.

A composition in accordance with the present invention, alone or in further combination with one or more other active agents, can be made into a formulation suitable for parenteral administration, preferably intraperitoneal administration. Such a formulation can include aqueous and nonaqueous, isotonic sterile injection

solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and nonaqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. The formulations can be presented in unit dose or multi-dose sealed containers, such as ampules and vials, and can be stored in a freeze-dried (lyophilized) condition requiring only the addition of the sterile liquid carrier, for example, water, for injections, immediately prior to use. Extemporaneously injectable solutions and suspensions can be prepared from sterile powders, granules, and tablets, as described herein.

A formulation suitable for oral administration can consist of liquid solutions, such as an effective amount of the compound dissolved in diluents, such as water, saline, or fruit juice; capsules, sachets or tablets, each containing a predetermined amount of the active ingredient, as solid or granules; solutions or suspensions in an aqueous liquid; and oil-in-water emulsions or water-in-oil emulsions. Tablet forms can include one or more of lactose, mannitol, corn starch, potato starch, microcrystalline cellulose, acacia, gelatin, colloidal silicon dioxide, croscarmellose sodium, talc, magnesium stearate, stearic acid, and other excipients, colorants, diluents, buffering agents, moistening agents, preservatives, flavoring agents, and pharmacologically compatible carriers.

Similarly, a formulation suitable for oral administration can include lozenge forms, which can comprise the active ingredient in a flavor, usually sucrose and acacia or tragacanth; pastilles comprising the active ingredient in an inert base, such as gelatin and glycerin, or sucrose and acacia; and mouthwashes comprising the

active ingredient in a suitable liquid carrier; as well as creams, emulsions, gels, and the like containing, in addition to the active ingredient, such carriers as are known in the art.

An aerosol formulation suitable for administration via inhalation also can be made. The aerosol formulation can be placed into a pressurized acceptable propellant, such as dichlorodifluoromethane, propane, nitrogen, and the like.

A formulation suitable for topical application can be in the form of creams, ointments, or lotions.

A formulation for rectal administration can be presented as a suppository with a suitable base comprising, for example, cocoa butter or a salicylate. A formulation suitable for vaginal administration can be presented as a pessary, tampon, cream, gel, paste, foam, or spray formula containing, in addition to the active ingredient, such carriers as are known in the art to be appropriate.

Important general considerations for design of delivery systems and compositions, and for routes of administration, for polypeptide drugs also apply (Eppstein, CRC Crit. Rev. Therapeutic Drug Carrier Systems 5, 99-139, 1988; Siddiqui et al., CRC Crit. Rev. Therapeutic Drug Carrier Systems 3, 195-208, 1987); Banga et al., Int. J. Pharmaceutics 48, 15-50, 1988; Sanders, Eur. J. Drug Metab. Pharmacokinetics 15, 95-102, 1990; Verhoef, Eur. J. Drug Metab. Pharmacokinetics 15, 83-93, 1990). The appropriate delivery system for a given polypeptide will depend upon its particular nature, the particular clinical application, and the site of drug action. As with any protein drug, oral delivery will likely present special problems, due primarily to instability in the gastrointestinal tract and poor absorption and bioavailability of intact, bioactive drug therefrom. Therefore, especially in the case of oral delivery, but also possibly in conjunction with other routes of delivery, it will be necessary to use an

absorption-enhancing agent in combination with a given polypeptide. A wide variety of absorption-enhancing agents have been investigated and/or applied in combination with protein drugs for oral delivery and for delivery by other routes (Verhoef, 1990, supra; van Hoogdalem, Pharmac. Ther. 44, 407-43, 1989; Davis, J. Pharm. Pharmacol. 44(Suppl. 1), 186-90, 1992). Most commonly, typical enhancers fall into the general categories of (a) chelators, such as EDTA, salicylates, and N-acyl derivatives of collagen, (b) surfactants, such as lauryl sulfate and polyoxyethylene-9-lauryl ether, (c) bile salts, such as glycholate and taurocholate, and derivatives, such as taurodihydrofusidate, (d) fatty acids, such as oleic acid and capric acid, and their derivatives, such as acylcarnitines, monoglycerides, and diglycerides, (e) non-surfactants, such as unsaturated cyclic ureas, (f) saponins, (g) cyclodextrins, and (h) phospholipids.

Other approaches to enhancing oral delivery of protein drugs can include the aforementioned chemical modifications to enhance stability to gastrointestinal enzymes and/or increased lipophilicity. Alternatively, the protein drug can be administered in combination with other drugs or substances that directly inhibit proteases and/or other potential sources of enzymatic degradation of proteins. Yet another alternative approach to prevent or delay gastrointestinal absorption of protein drugs is to incorporate them into a delivery system that is designed to protect the protein from contact with the proteolytic enzymes in the intestinal lumen and to release the intact protein only upon reaching an area favorable for its absorption. A more specific example of this strategy is the use of biodegradable microcapsules or microspheres, both to protect vulnerable drugs from degradation, as well as to effect a prolonged release of active drug (Deasy, in Microencapsulation and Related Processes, Swarbrick, ed.,

Marcell Dekker, Inc.: New York, 1984, pp. 1-60, 88-89, 208-11). Microcapsules also can provide a useful way to effect a prolonged delivery of a protein drug after injection (Maulding, J. Controlled Release 6, 167-76, 1987).

The dose administered to an animal, such as a mammal, particularly a human, in the context of the present invention should be sufficient to effect a therapeutic or prophylactic response in the individual over a reasonable time frame. The dose will be determined by the particular polypeptide, nucleic acid, antibody, or anti-antibody administered, the severity of any existing disease state, as well as the body weight and age of the individual. The size of the dose also will be determined by the existence of any adverse side effects that may accompany the use of the particular polypeptide, nucleic acid, antibody or anti-antibody employed. It is always desirable, whenever possible, to keep adverse side effects to a minimum.

The dosage can be in unit dosage form, such as a tablet or capsule. The term "unit dosage form" as used herein refers to physically discrete units suitable as unitary dosages for human and animal subjects, each unit containing a predetermined quantity of a vector, alone or in combination with other active agents, calculated in an amount sufficient to produce the desired effect in association with a pharmaceutically acceptable diluent, carrier, or vehicle. The specifications for the unit dosage forms of the present invention depend on the particular embodiment employed and the effect to be achieved, as well as the pharmacodynamics associated with each polypeptide, nucleic acid or anti-antibody in the host. The dose administered should be an "HIV infection inhibiting amount" of an above-described polypeptide or nucleic acid or an "immune response-inducing effective amount" of an above-described polypeptide, an above-described nucleic acid, or an antibody as appropriate.

Another composition provided by the present invention is a composition comprising a solid support matrix to which is attached an above-described polypeptide, or an anti-antibody to an above-described polypeptide. The solid matrix can comprise other functional reagents including, for example, polyethylene glycol, dextran, albumin and the like, whose intended effector functions may include one or more of the following: to improve stability of the conjugate; to increase the half-life of the conjugate; to increase resistance of the conjugate to proteolysis; to decrease the immunogenicity of the conjugate; to provide a means to attach or immobilize a functional polypeptide or anti-antibody onto a solid support matrix (e.g., see, for example, Harris, in <u>Poly(Ethylene Glycol) Chemistry: Biotechnical and Biomedical Applications</u>, Harris, ed., Plenum Press: New York (1992), pp. 1-14). Conjugates furthermore may comprise a polypeptide or anti-antibody coupled to an effector molecule, each of which, optionally, may have different functions (e.g., such as a toxin molecule (or an immunological reagent) and a polyethylene glycol (or dextran or albumin) molecule). Diverse applications and uses of functional proteins and polypeptides, attached to or immobilized on a solid support matrix, are exemplified more specifically for poly(ethylene glycol) conjugated proteins or peptides in a review by Holmberg et al. (In <u>Poly(Ethylene Glycol) Chemistry: Biotechnical and Biomedical Applications,</u> Harris, ed., Plenum Press: New York, 1992, pp. 303-324).

In addition, the present invention provides a method of removing HIV from a bodily fluid of an animal. The method comprises extracorporeally contacting the bodily fluid of the animal with a solid-support matrix to which is attached an above-described polypeptide or an anti-antibody to an above-described polypeptide. Alternatively, the bodily fluid can be contacted with the polypeptide or anti-

antibody in solution and then the solution can be contacted with a solid support matrix to which is attached a means to remove the polypeptide or anti-antibody to which is bound HIV gp120 from the bodily fluid.

Methods of attaching an herein-described polypeptide, or an anti-antibody to a solid support matrix are known in the art. "Attached" is used herein to refer to attachment to (or coupling to) and immobilization in or on a solid support matrix. See, for example, Harris, in Poly(Ethylene Glycol) Chemistry: Biotechnical and Biomedical Applications, Harris, ed., Plenum Press: New York (1992), pp. 1-14) and international patent application WO 91/02714 (Saxinger). Diverse applications and uses of functional polypeptides attached to or immobilized on a solid support matrix are exemplified more specifically for poly(ethylene glycol) conjugated proteins or peptides in a review by Holmberg et al. (In Poly(Ethylene Glycol) Chemistry: Biotechnical and Biomedical Applications, Harris, ed., Plenum Press: New York, 1992, pp. 303-324).

The present invention also provides a method of making an antibody that binds to gp120 of HIV under physiological conditions. The method comprises labeling an embodiment of the present inventive compound to obtain a labeled compound. Labeling compounds are within the skill of the ordinary artisan. For example, the present inventive compound can be labeled with radioactive atom, such as ^{125}I in the same or a similar manner as was performed in the examples provided below. Alternatively, an enzyme, such as horseradish peroxidase, can be attached to or incorporated into the present inventive compound. Then by exposing a chromogenic or photogenic compound to the compound, a signal indicative of the presence and quantity of the compound present can be generated. In another alternative, a polyhistidinyl moiety can be attached to, or incorporated into, the present inventive moiety so that the present

inventive compound will react with high affinity to transition metal ions such as nickel, copper, or zinc ions; this reaction can be used as the basis to quantify the amount of the present inventive compound present at a particular location. In yet another alternative, the present inventive compound can be used as antigen to a standard antibody that specifically recognizes an antigenic epitope of the present inventive compound. As is well-known, the standard antibody can itself be labeled or used in conjunction with an additional antibody that is labeled with an enzyme, radioisotope, or other suitable means. The skilled artisan will recognize that there is a plethora of other suitable means and methods to label the present inventive compound.

This present inventive method of making an antibody that binds to a gp120 envelope protein of HIV further comprises providing a library of synthetic peptides. The library consists of a multiplicity of synthetically-produced polypeptides that are homologous, and preferably essentially identical (i.e., having the same primary amino acid residue sequence, ignoring blocking groups, phosphorylation of serinyl, threoninyl, and tyrosinyl residues, hydroxylation of prolinyl residues, and the like) or identical, to a continuous region of an HIV gp120 envelope protein. The polypeptides of the library can be any suitable length. While larger regions allow faster scanning and tend to preserve non-linear epitopes, shorter length polypeptides allow more sensitive screening of the primary sequence of the gp120 protein. However, polypeptides that are too short can lose essential secondary structure or cleave reactive sites into one or more pieces. Preferably, a mixture of short and long polypeptides are incorporated into the library, however, the library can consist of polypeptides of a single length (measured in amino acid residues). For the sake of

convenience the library can be split into multiple parts, and screened by parts. Typically, the polypeptides of the library will be between about 6 and about 45 amino acid residues in length.

Typically, the library will comprise a series of polypeptides each having an identical sequence to that of gp120 but having an amino-terminus a particular number of amino acids downstream of the amino-terminus of the prior polypeptide (see, examples section below). The distance, measured in amino acid residues, is referred to as the offset. Preferably, libraries that are characterized by the existence of an offset, the offset is not greater than the product of length of the longest polypeptide measured in amino acid residues and 1.5, preferably 1.0, and more preferably 0.5. The library can be alternatively characterized by the existence of an offset not greater than 30, preferably 15, and more preferably 4.

Each polypeptide of the library is substantially isolated from every other polypeptide of said library and is located in a known position. For example, each polypeptide can be bound to a solid support and that is in a vessel or that can be placed in a vessel. The vessel preferably enables each polypeptide to be covered in a liquid that does not contact any other oligonucleotide of the library. By way of example, each polypeptide can be bound to a bead that is placed in a vessel (or tube) or can be bound to the well of a multi-well assay plate. Alternatively, an array of polypeptides can be fashioned, for example on a microchip device (as is presently used in some DNA sequencing devices and methods), and the entire array can be bathed in a single solution.

Each polypeptide is then individually contacted with the labeled compound such that a portion of the labeled compound can bind with the polypeptide of the library. In this way, a bound population of each labeled compound of

the present invention and an unbound population of the labeled compound is generated. The phrase individually contacted means that each polypeptide has the opportunity to bind with the labeled compound and the quantity of labeled compound bound by each can be determined.

The method then comprises removing substantially all of the unbound labeled compound from the position occupied by each polypeptide. That is, the solution comprising the labeled compound is separated from the polypeptides of the library and the bound population of the labeled compound. This can be done by any suitable method, e.g., by aspiration and one or more washing steps comprising adding a quantity of liquid sufficient to cover all the surfaces that were contacted by the labeled compound and aspirating away substantially all of the wash liquid.

The amount of labeled compound that remains co-localized with each polypeptide of the library is then measured to determine the quantity of labeled compound bound by each polypeptide. The amount of the present inventive compound bound by each polypeptide can be directly evaluated to identify a portion of the HIV gp120 envelope protein that binds to an (HIV)-receptor selected from the group consisting of CCR5, CXCR4, STRL33, and CD4. This information is then used to identify and provide an immunizing compound. The immunizing compound comprises a polypeptide comprising an amino acid sequence that is homologous to, or preferably is essentially identical to, or identical to, the portion of the HIV-1 gp120 envelope protein that binds with CD4, CCR5, CXCR4, and/or STRL33. The immunizing protein can be provided by processing gp120, e.g., proteolytically digesting gp120 that has been isolated from a preparation of HIV-1. Preferably, however, the immunizing compound is prepared synthetically, or by genetic engineering, or by a combination of genetic engineering and synthetic methods. The immunizing compound

can comprise a pharmaceutically acceptable substituent, can be encoded by a nucleic acid that can be expressed in a cell, can be mixed with a carrier, and is an inventive aspect of the present invention.

An immunogenic quantity of the immunizing compound is then inserted into an animal (e.g., a human, or a rodent, a canine, a feline, or a ruminant) in a manner consistent with the discussion of a method of raising an antibody to the present inventive compounds that are homologous to portions of CCR5, CXCR4, STRL33, and CD4, above. The insertion of the immunizing compound causes the inoculated animal to produce an antibody that binds with said portion of the HIV gp120 envelope protein. Thus the present invention also provides an antibody that binds to an HIV gp120 envelope protein, as well as an antigen binding protein comprising one or more complementarity determining regions of the antibody (e.g., a Fab, a $Fab_{2'}$, an Fv, a single-chain antibody, a diabody, and humanized variants of all of the above, all of which are within the skill in the art).

The antibody or variant thereof is preferably useful in detecting or diagnosing the presence of HIV gp120 envelope protein, and thus HIV, in an animal. The antibody is also preferably prevents or attenuates infection of an animal exposed to HIV, to whom an effective quantity of the antibody or a variant thereof, has been administered or produced in response to inoculation with the immunizing compound. The antibody preferably also is useful in treating or preventing (i.e., inhibiting) HIV infection in an animal to whom a suitable dose has been administered or in which a suitable quantity of antibody has been produced. The antibody is also useful in the study of HIV infection of mammalian cells, the host range specificities of HIV infection, and preferably, the mechanism by which antibodies neutralize infectious viruses.

EXAMPLES

The following examples further illustrate the present invention but, of course, should not be construed as limiting the scope of the claimed invention in any way.

Synthetic peptide arrays were constructed in 96-well microtiter plates in accordance with the method set forth in WO 91/02714 (Saxinger), and used to test the binding of HIV-1_{LAI} envelope gp120 that had been labeled with radioactive iodine (radiolabeling by standard methods). After incubating the radiolabeled gp120 in a well with each synthetic peptide, a washing step was performed to remove unbound label, and the relative level of radioactivity remaining in each well of the plate was evaluated to determine the relative affinity of each peptide for the gp120. The synthesis of the peptides and the quantity of binding between the synthetic peptides and the gp120 were found to be suitably reproducible, precise, and sensitive. Initial screening of the entire primary sequence of the chemokine and CD4 receptor molecules was taken 18 amino acid residues at a time.

The authenticity of the binding signals generated by this technique has been repeatedly demonstrated by showing that antibodies to CCR5 and CXCR4 are able to inhibit the binding of radiolabeled gp120 to the polypeptides derived from CCR5 and CXCR4 that show a high affinity for binding with gp120. Additionally, the accuracy of the binding assay used hereinbelow is demonstrated by Example 7.

<u>Example 1</u>

This example identifies segments of the CCR5 co-receptor that bind with gp120.

The first column in the table below indicates the number of the amino acid in the wild-type CCR5 receptor. The second column explicitly identifies the peptide

sequence. The third column indicates the radioactive counts recorded in twenty minutes (i.e., the cpm x 20) after the background or non-specific counts had been subtracted. The fourth column contains an X in each row for which the listed polypeptide bound with high affinity to gp120. The fifth and final column contains an X in each row wherein the listed sequence binds with substantial affinity but is weak in comparison to other samples, particularly adjacent samples.

SEQ	SEG	PEPTIDE	Counts per 20' Average-background	Peak Activity	Non-Peak Activity	SEQ ID NO:
		empty (control)	7			
1--18		MDYQVSSPIYDINYYTSE	735	X		31
5--22		VSSPIYDINYYTSEPCQK	383		x	32
9--26		IYDINYYTSEPCQKINVK	228		x	33
13-30		NYYTSEPCQKINVKQIAA	6			34
17-34		SEPCQKINVKQIAARLLP	-44			35
21-38		QKINVKQIAARLLPPLYS	20			36
25-42		VKQIAARLLPPLYSLVFI	18			37
29-46		AARLLPPLYSLVFIFGFV	33			38
33-50		LPPLYSLVFIFGFVGNML	705	X		39
37-54		YSLVFIFGFVGNMLVILI	347		x	40
41-58		FIFGFVGNMLVILILINC	343		x	41
45-62		FVGNMLVILILINCKRLK	62			42
49-66		MLVILILINCKRLKSMTD	84			43
53-70		LILINCKRLKSMTDIYLL	2			44
57-74		NCKRLKSMTDIYLLNLAI	25			45
61-78		LKSMTDIYLLNLAISDLF	210			46
65-82		TDIYLLNLAISDLFFLLT	38			47
69-86		LLNLAISDLFFLLTVPFW	144			48
73-90		AISDLFFLLTVPFWAHYA	41			49
77-94		LFFLLTVPFWAHYAAAQW	173			50
81-98		LTVPFWAHYAAAQWDFGN	306			51
85-		FWAHYAAAQWDFGNTMCQ	212			52
89-		YAAAQWDFGNTMCQLLTG	494		x	53
93-		QWDFGNTMCQLLTGLYFI	1019	X		54
97-		GNTMCQLLTGLYFIGFFS	941	X		55
101-		CQLLTGLYFIGFFSGIFF	489		x	56
105-		TGLYFIGFFSGIFFIILL	80			57
109-		FIGFFSGIFFIILLTIDR	76			58
113-		FSGIFFIILLTIDRYLAV	83			59
117-		FFIILLTIDRYLAVVHAV	77			60
121-		LLTIDRYLAVVHAVFALK	31			61
125-		DRYLAVVHAVFALKARTV	62			62
129-		AVVHAVFALKARTVTFGV	34			63
133-		AVFALKARTVTFGVVTSV	63			64
137-		LKARTVTFGVVTSVITWV	74			65
141-		TVTFGVVTSVITWVVAVF	-25			66
145-		GVVTSVITWVVAVFASLP	69			67
149-		SVITWVVAVFASLPGIIF	46			68
153-		WVVAVFASLPGIIFTRSQ	87			69
157-		VFASLPGIIFTRSQKEGL	54			70
161-		LPGIIFTRSQKEGLHYTC	118			71
165-		IFTRSQKEGLHYTCSSHF	98			72

169-	SQKEGLHYTCSSHFPYSQ	304		x	73
173-	GLHYTCSSHFPYSQYQFW	301		x	74
177-	TCSSHFPYSQYQFWKNFQ	367		x	75
181-	HFPYSQYQFWKNFQTLKI	1008		x	76
185-	SQYQFWKNFQTLKIVILG	1572	X		77
189-	FWKNFQTLKIVILGLVLP	40			78
193-	FQTLKIVILGLVLPLLVM	45			79
197-	KIVILGLVLPLLVMVICY	65			80
201-	LGLVLPLLVMVICYSGIL	180			81
205-	LPLLVMVICYSGILKTLL	68			82
209-	VMVICYSGILKTLLRCRN	-8			83
213-	CYSGILKTLLRCRNEKKR	70			84
217-	ILKTLLRCRNEKKRHRAV	19			85
221-	LLRCRNEKKRHRAVRLIF	102			86
225-	RNEKKRHRAVRLIFTIMI	23			87
229-	KRHRAVRLIFTIMIVYFL	36			88
233-	AVRLIFTIMIVYFLFWAP	62			89
237-	IFTIMIVYFLFWAPYNIV	121			90
241-	MIVYFLFWAPYNIVLLLN	214			91
245-	FLFWAPYNIVLLLNTFQE	616		x	92
249-	APYNIVLLLNTFQEFFGL	1962	X		93
253-	IVLLLNTFQEFFGLNNCS	2134	X		94
257-	LNTFQEFFGLNNCSSSNR	293		x	95
261-	QEFFGLNNCSSSNRLDQA	63			96
265-	GLNNCSSSNRLDQAMQVT	-31			97
269-	CSSSNRLDQAMQVTETLG	90			98
273-	NRLDQAMQVTETLGMTHC	10			99
277-	QAMQVTETLGMTHCCINP	81			100
281-	VTETLGMTHCCINPIIYA	15			101
285-	LGMTHCCINPIIYAFVGE	282		x	102
289-	HCCINPIIYAFVGEKFRN	200		x	103
293-	NPIIYAFVGEKFRNYLLV	162		x	104
297-	YAFVGEKFRNYLLVFFQK	596	X		105
301-	GEKFRNYLLVFFQKHIAK	69			106
305-	RNYLLVFFQKHIAKRFCK	65			107
309-	LVFFQKHIAKRFCKCCSI	76			108
313-	QKHIAKRFCKCCSIFQQE	23			109
317-	AKRFCKCCSIFQQEAPER	64			110
321-	CKCCSIFQQEAPERASSV	53			111
325-	SIFQQEAPERASSVYTRS	100			112
329-	QEAPERASSVYTRSTGEQ	84			113
333-	ERASSVYTRSTGEQEISV	84			114
337-	SVYTRSTGEQEISVGL	47			115

These data indicate that, in addition to polypeptide sequences derived from positions 1-18 of the CCR5 receptor, the polypeptide sequences LPPLYSLVFIFGFVGNML (SEQ ID NO: 11), QWDFGNTMCQLLTGLYFIGFFS (SEQ ID NO: 12),

SQYQFWKNFQTLKIVILG (SEQ ID NO: 13), APYNIVLLLNTFQEFFGLNNCS (SEQ ID NO: 14), and YAFVGEKFRNYLLVFFQK (SEQ ID NO: 15) comprise multiple subsequences, each which is capable of binding to HIV-1 envelope gp120.

Example 2

This example identifies segments of the CXCR4 co-receptor that bind with gp120.

The first column in the table below indicates the number of the amino acid in the wild-type CXCR4 receptor. The second column explicitly identifies the peptide sequence. The third and fourth columns indicate the radioactive counts recorded in twenty minutes (i.e., the cpm x 20) after the background or non-specific counts had been subtracted. The fifth column contains an X in each row for which the listed polypeptide bound with high affinity to gp120. The sixth and final column contains an X in each row wherein the listed sequence binds with substantial affinity but is weak in comparison to other samples, particularly adjacent samples.

SEQ SEG	PEPTIDE			Major Activity Peak	Minor Activity Peak	SEQ ID NO:
	empty (control)	412	0			
1—18	MEGISIYTSDNYTEEMGS	3003	2591	X		116
5--22	SIYTSDNYTEEMGSGDYD	483	71			117
9--26	SDNYTEEMGSGDYDSMKE	455	43			118
13-30	TEEMGSGDYDSMKEPCFR	453	41			119
17-34	GSGDYDSMKEPCFREENA	384	-28			120
21-38	YDSMKEPCFREENANFNK	465	53			121
25-42	KEPCFREENANFNKIFLP	664	252			122
29-46	FREENANFNKIFLPTIYS	463	51			123
33-50	NANFNKIFLPTIYSIIFL	585	173			124
37-54	NKIFLPTIYSIIFLTGIV	550	138			125
41-58	LPTIYSIIFLTGIVGNGL	530	118			126
45-62	YSIIFLTGIVGNGLVILV	535	123			127
49-66	FLTGIVGNGLVILVMGYQ	658	246			128
53-70	IVGNGLVILVMGYQKKLR	650	238			129
57-74	GLVILVMGYQKKLRSMTD	569	157			130
61-78	LVMGYQKKLRSMTDKYRL	517	105			131
65-82	YQKKLRSMTDKYRLHLSV	511	99			132
69-86	LRSMTDKYRLHLSVADLL	572	160			133
73-90	TDKYRLHLSVADLLFVIT	504	92			134
77-94	RLHLSVADLLFVITLPFW	548	136			135
81-98	SVADLLFVITLPFWAVDA	665	253			136
85-102	LLFVITLPFWAVDAVANW	475	63			137
89-106	ITLPFWAVDAVANWYFGN	542	130			138
93-110	FWAVDAVANWYFGNFLCK	478	66			139
97-114	DAVANWYFGNFLCKAVHV	524	112			140
101-118	NWYFGNFLCKAVHVIYTV	508	96			141
105-122	GNFLCKAVHVIYTVNLYS	643	231			142
109-126	CKAVHVIYTVNLYSSVLI	655	243			143
113-130	HVIYTVNLYSSVLILAFI	530	118			144
117-134	TVNLYSSVLILAFISLDR	654	242			145
121-138	YSSVLILAFISLDRYLAI	569	157			146
125-142	LILAFISLDRYLAIVHAT	519	107			147
129-146	FISLDRYLAIVHATNSQR	503	91			148
133-150	DRYLAIVHATNSQRPRKL	580	168			149
137-154	AIVHATNSQRPRKLLAEK	485	73			150
141-158	ATNSQRPRKLLAEKVVYV	490	78			151
145-162	QRPRKLLAEKVVYVGVWI	539	127			152
149-166	KLLAEKVVYVGVWIPALL	501	89			153
153-170	EKVVYVGVWIPALLLTIP	559	147			154
157-174	YVGVWIPALLLTIPDFIF	536	124			155
161-178	WIPALLLTIPDFIFANVS	594	182			156
165-182	LLLTIPDFIFANVSEADD	1418	1006	X		157
169-186	IPDFIFANVSEADDRYIC	850	438		x	158
173-190	IFANVSEADDRYICDRFY	679	267			159
177-194	VSEADDRYICDRFYPNDL	569	157			160
181-198	DDRYICDRFYPNDLWVVV	537	125			161

185-202	ICDRFYPNDLWVVVFQFQ	718	306			162
189-206	FYPNDLWVVVFQFQHIMV	828	416		x	163
193-210	DLWVVVFQFQHIMVGLIL	834	422	X		164
197-214	VVFQFQHIMVGLILPGIV	1001	589		x	165
201-218	FQHIMVGLILPGIVILSC	582	170			166
205-222	MVGLILPGIVILSCYCII	579	167			167
209-226	ILPGIVILSCYCIIISKL	604	192			168
213-230	IVILSCYCIIISKLSHSK	689	277			169
217-234	SCYCIIISKLSHSKGHQK	671	259			170
221-238	IIISKLSHSKGHQKRKAL	569	157			171
225-242	KLSHSKGHQKRKALKTTV	542	130			172
229-246	SKGHQKRKALKTTVILIL	552	140			173
233-250	QKRKALKTTVILILAFFA	695	283			174
237-254	ALKTTVILILAFFACWLP	673	261			175
241-258	TVILILAFFACWLPYYIG	735	323			176
245-262	ILAFFACWLPYYIGISID	596	184			177
249-266	FACWLPYYIGISIDSFIL	614	202			178
253-270	LPYYIGISIDSFILLEII	851	439			179
257-274	IGISIDSFILLEIIKQGC	1146	734		x	180
261-278	IDSFILLEIIKQGCEFEN	3884	3472	X		181
265-282	ILLEIIKQGCEFENTVHK	529	117			182
269-286	IIKQGCEFENTVHKWISI	518	106			183
273-290	GCEFENTVHKWISITEAL	676	264			184
277-294	ENTVHKWISITEALAFFH	727	315			185
281-298	HKWISITEALAFFHCCLN	575	163			186
285-302	SITEALAFFHCCLNPILY	600	188			187
289-306	ALAFFHCCLNPILYAFLG	593	181			188
293-310	FHCCLNPILYAFLGAKFK	535	123			189
297-314	LNPILYAFLGAKFKTSAQ	686	274			190
301-318	LYAFLGAKFKTSAQHALT	568	156			191
305-322	LGAKFKTSAQHALTSVSR	612	200			192
309-326	FKTSAQHALTSVSRGSSL	585	173			193
313-330	AQHALTSVSRGSSLKILS	559	147			194
317-334	LTSVSRGSSLKILSKGKR	595	183			195
321-338	SRGSSLKILSKGKRGGHS	581	169			196
325-342	SLKILSKGKRGGHSSVST	697	285			197
329-346	LSKGKRGGHSSVSTESES	597	185			198
333-350	KRGGHSSVSTESESSSFH	579	167			199
337-352	HSSVSTESESSSFHSS	515	103			200

These data indicate that, in addition to polypeptide sequences derived from positions 1-18 of the CXCR4 receptor, the polypeptide sequences LLLTIPDFIFANVSEADD (SEQ ID NO: 16)(165-182), VVFQFQHIMVGLILPGIV (SEQ ID NO: 17) (197-214), and IDSFILLEIIKQGCEFEN (SEQ ID NO: 18) (261-278) comprise multiple subsequences, which is capable of binding to HIV-1 envelope gp120.

Example 3

This example identifies segments of the STRL33 co-receptor that bind with gp120.

The first column in the table below indicates the number of the amino acid in the wild-type STRL33 receptor. The second column explicitly identifies the peptide sequence. The third and fourth columns indicate the radioactive counts recorded in twenty minutes (i.e., the cpm x 20) after the background or non-specific counts had been subtracted. The fifth column contains an X in each row for which the listed polypeptide bound with high affinity to gp120. The sixth and final column contains an X in each row wherein the listed sequence binds with substantial affinity but is weak in comparison to other samples, particularly adjacent samples.

SEQ SEG	PEPTIDE			Major Activity Peak	Minor Activity Peak	SEQ ID NO:
	empty (control)	-34.5	34.5			
1--18	MAEHDYHEDYGFSSFNDS	1178.5	1320.5		X	201
5--22	DYHEDYGFSSFNDSSQEE	3357.5	3689.5		X	202
9--26	DYGFSSFNDSSQEEHQAF	8579.5	8909.5	X		203
13-30	SSFNDSSQEEHQAFLQFS	2689.5	2757.5		X	204
17-34	DSSQEEHQAFLQFSKVFL	869.5	2152.5		X	205
21-38	EEHQAFLQFSKVFLPCMY	2316.5	1819.5		X	206
25-42	AFLQFSKVFLPCMYLVVF	1421.5	1359.5		X	207
29-46	FSKVFLPCMYLVVFVCGL	534.5	633.5			208
33-50	FLPCMYLVVFVCGLVGNS	605.5	372.5			209
37-54	MYLVVFVCGLVGNSLVLV	168.5	235.5			210
41-58	VFVCGLVGNSLVLVISIF	570.5	284.5			211
45-62	GLVGNSLVLVISIFYHKL	164.5	95.5			212
49-66	NSLVLVISIFYHKLQSLT	1255.5	1378.5		X	213
53-70	LVISIFYHKLQSLTDVFL	1620.5	1780.5	X		214
57-74	IFYHKLQSLTDVFLVNLP	1275.5	1256.5		X	215
61-78	KLQSLTDVFLVNLPLADL	412.5	348.5			216
65-82	LTDVFLVNLPLADLVFVC	233.5	336.5			217
69-86	FLVNLPLADLVFVCTLPF	70.5	51.5			218
73-90	LPLADLVFVCTLPFWAYA	557.5	960.5		X	219
77-94	DLVFVCTLPFWAYAGIHE	1116.5	1063.5		X	220
81-98	VCTLPFWAYAGIHEWVFG	1819.5	1754.5		X	221
85-102	PFWAYAGIHEWVFGQVMC	7262.5	7537.5	X		222
89-106	YAGIHEWVFGQVMCKSLL	5911.5	6245.5		X	223
93-110	HEWVFGQVMCKSLLGIYT	3391.5	3466.5		X	224
97-114	FGQVMCKSLLGIYTINFY	1257.5	1354.5		X	225
101-118	MCKSLLGIYTINFYTSML	1505.5	1283.5			226
105-122	LLGIYTINFYTSMLILTC	499.5	408.5			227
109-126	YTINFYTSMLILTCITVD	351.5	510.5			228
113-130	FYTSMLILTCITVDRFIV	744.5	907.5			229
117-134	MLILTCITVDRFIVVVKA	298.5	228.5			230
121-138	TCITVDRFIVVVKATKAY	89.5	346.5			231
125-142	VDRFIVVVKATKAYNQQA	103.5	53.5			232
129-146	IVVVKATKAYNQQAKRMT	166.5	43.5			233
133-150	KATKAYNQQAKRMTWGKV	701.5	568.5			234
137-154	AYNQQAKRMTWGKVTSLL	55.5	4.5			235
141-158	QAKRMTWGKVTSLLIWVI	-71.5	-31.5			236
145-162	MTWGKVTSLLIWVISLLV	-0.5	-26.5			237
149-166	KVTSLLIWVISLLVSLPQ	-39.5	-118.5			238
153-170	LLIWVISLLVSLPQIIYG	42.5	75.5			239
157-174	VISLLVSLPQIIYGNVFN	-60.5	-127.5			240
161-178	LVSLPQIIYGNVFNLDKL	91.5	-15.5			241
165-182	PQIIYGNVFNLDKLICGY	-18.5	-37.5			242
169-186	YGNVFNLDKLICGYHDEA	-41.5	-20.5			243
173-190	FNLDKLICGYHDEAISTV	1072.5	1078.5		X	244
177-194	KLICGYHDEAISTVVLAT	1363.5	1604.5		X	245

181-198	GYHDEAISTVVLATQMTL	754.5	1181.5		X	246
185-202	EAISTVVLATQMTLGFFL	3973.5	3745.5	X		247
189-206	TVVLATQMTLGFFLPLLT	2327.5	2389.5		X	248
193-210	ATQMTLGFFLPLLTMIVC	2365.5	2444.5		X	249
197-214	TLGFFLPLLTMIVCYSVI	2387.5	479.5			250
201-218	FLPLLTMIVCYSVIIKTL	1270.5	1195.5		X	251
205-222	LTMIVCYSVIIKTLLHAG	2787.5	2654.5	X		252
209-226	VCYSVIIKTLLHAGGFQK	1334.5	1143.5		X	253
213-230	VIIKTLLHAGGFQKHRSL	961.5	682.5			254
217-234	TLLHAGGFQKHRSLKIIF	1041.5	999.5			255
221-238	AGGFQKHRSLKIIFLVMA	340.5	260.5			256
225-242	QKHRSLKIIFLVMAVFLL	810.5	814.5			257
229-246	SLKIIFLVMAVFLLTQMP	612.5	853.5			258
233-250	IFLVMAVFLLTQMPFNLM	386.5	772.5			259
237-254	MAVFLLTQMPFNLMKFIR	2263.5	2842.5	X		260
241-258	LLTQMPFNLMKFIRSTHW	2513.5	3154.5	X		261
245-262	MPFNLMKFIRSTHWEYYA	2171.5	2182.5		X	262
249-266	LMKFIRSTHWEYYAMTSF	934.5	949.5			263
253-270	IRSTHWEYYAMTSFHYTI	1571.5	1807.5		X	264
257-274	HWEYYAMTSFHYTIMVTE	2040.5	3065.5	X		265
261-278	YAMTSFHYTIMVTEAIAY	2688.5	2359.5		X	266
265-282	SFHYTIMVTEAIAYLRAC	761.5	1033.5			267
269-286	TIMVTEAIAYLRACLNPV	140.5	272.5			268
273-290	TEAIAYLRACLNPVLYAF	604.5	480.5			269
277-294	AYLRACLNPVLYAFVSLK	1802.5	1849.5		X	270
281-298	ACLNPVLYAFVSLKFRKN	4173.5	4515.5	X		271
285-302	PVLYAFVSLKFRKNFWKL	1859.5	2147.5		X	272
289-306	AFVSLKFRKNFWKLVKDI	808.5	1040.5			273
293-310	LKFRKNFWKLVKDIGCLP	920.5	957.5			274
297-314	KNFWKLVKDIGCLPYLGV	143.5	82.5			275
301-318	KLVKDIGCLPYLGVSHQW	-2.5	27.5			276
305-322	DIGCLPYLGVSHQWKSSE	17.5	78.5			277
309-326	LPYLGVSHQWKSSEDNSK	111.5	122.5			278
313-330	GVSHQWKSSEDNSKTFSA	208.5	306.5			279
317-334	QWKSSEDNSKTFSASHNV	464.5	533.5			280
321-338	SEDNSKTFSASHNVEATS	524.5	434.5			281
325-342	SKTFSASHNVEATSMFQL	1524.5	1239.5	X		282

These data indicate that, in addition to polypeptide sequences derived from positions 9-26 of the STRL33 receptor, the polypeptide sequences LVISIFYHKLQSLTDVFL (SEQ ID NO: 19) (53-70), PFWAYAGIHEWVFGQVMC (SEQ ID NO: 20)(85-102), EAISTVVLATQMTLGFFL (SEQ ID NO: 21)(185-202), LTMIVCYSVIIKTLLHAG (SEQ ID NO: 22)(205-222), MAVFLLTQMPFNLMKFIRSTHW (SEQ ID NO: 23)(237-258), HWEYYAMTSFHYTIMVTE (SEQ ID NO: 24)(257-274), ACLNPVLYAFVSLKFRKN (SEQ ID NO: 25)(281-298) and

SKTFSASHNVEATSMFQL (SEQ ID NO: 26)(325-342) comprise multiple subsequences, which is capable of binding to HIV-1 envelope gp120.

Example 4

This example identifies segments of the human CD4 protein that bind with gp120.

The second column in the in the table below identifies the amino acid residue sequence of the polypeptide employed in the assay. The first column identifies the sequence coordinates of human CD4 that have an identical amino acid sequence. The third column indicates the number of radioactive decays (i.e., counts) that were counted, which is indicative of the affinity of the synthetic polypeptide for the gp120 protein. In the table below, polypeptides retaining more than 4,000 counts identify fragments that have a substantial capability to bind with gp120. Polypeptides retaining more than 6,000 counts have more substantial binding affinity. Polypeptides retaining at least about 10,000 counts have a substantial and strong capacity to bind to gp120. Of course, fragments corresponding to amino acid coordinates 101-121 and 106-126 have a substantial, strong, and dominant capacity to bind to gp120.

					SEQ ID NO:
B1	(1)	1-21	MNRGVPFRHLLLVLQLALLPA	3587	283
C1	(2)	6-26	PFRHLLLVLQLALLPAATQGK	4356	284
D1	(3)	11-31	LLVLQLALLPAATQGKKVVLG	1785	285
E1	(4)	16-36	LALLPAATQGKKVVLGKKGDT	1759	286
F1	(5)	21-41	AATQGKKVVLGKKGDTVELTC	1562	287
G1	(6)	26-46	KKVVLGKKGDTVELTCTASQK	1910	288
H1	(7)	31-51	GKKGDTVELTCTASQKKSIQF	1831	289
A2	(8)	36-56	TVELTCTASQKKSIQFHWKNS	1732	290
B2	(9)	41-61	CTASQKKSIQFHWKNSNQIKI	1717	291
C2	(10)	46-66	KKSIQFHWKNSNQIKILGNQG	2182	292
D2	(11)	51-71	FHWKNSNQIKILGNQGSFLTK	1835	293
E2	(12)	56-76	SNQIKILGNQGSFLTKGPSKL	1487	294
F2	(13)	61-81	ILGNQGSFLTKGPSKLNDRAD	1467	295
G2	(14)	66-86	GSFLTKGPSKLNDRADSRRSL	1844	296
H2	(15)	71-91	KGPSKLNDRADSRRSLWDQGN	1912	297
A3	(16)	76-96	LNDRADSRRSLWDQGNFPLII	1753	298

B3 (17)	81-101	DSRRSLWDQGNFPLIIKNLKI	2224	299
C3 (18)	86-106	LWDQGNFPLIIKNLKIEDSDT	3264	300
D3 (19)	91-111	NFPLIIKNLKIEDSDTYICEV	11646	301
E3 (20)	96-116	IKNLKIEDSDTYICEVEDQKE	8439	302
F3 (21)	101-121	IEDSDTYICEVEDQKEEVQLL	6803	303
G3 (22)	106-126	TYICEVEDQKEEVQLLVFGLT	44965	304
H3 (23)	111-131	VEDQKEEVQLLVFGLTANSDT	36249	305
A4 (24)	116-136	EEVQLLVFGLTANSDTHLLQG	14171	306
B4 (25)	121-141	LVFGLTANSDTHLLQGQSLTL	3683	307
C4 (26)	126-146	TANSDTHLLQGQSLTLTLESP	6114	308
D4 (27)	131-151	THLLQGQSLTLTLESPPGSSP	2552	309
E4 (28)	136-156	GQSLTLTLESPPGSSPSVQCR	1538	310
F4 (29)	141-161	LTLESPPGSSPSVQCRSPRGK	1476	311
G4 (30)	146-166	PPGSSPSVQCRSPRGKNIQGG	1496	312
H4 (31)	151-171	PSVQCRSPRGKNIQGGKTLSV	1400	313
A5 (32)	156-176	RSPRGKNIQGGKTLSVSQLEL	2066	314
B5 (33)	161-181	KNIQGGKTLSVSQLELQDSGT	3078	315
C5 (34)	166-186	GKTLSVSQLELQDSGTWTCTV	2618	316
D5 (35)	171-191	VSQLELQDSGTWTCTVLQNQK	3879	317
E5 (36)	176-196	LQDSGTWTCTVLQNQKKVEFK	2456	318
F5 (37)	181-201	TWTCTVLQNQKKVEFKIDIVV	4030	319
G5 (38)	186-206	VLQNQKKVEFKIDIVVLAFQK	9737	320
H5 (39)	191-211	KKVEFKIDIVVLAFQKASSIV	6313	321
A6 (40)	196-216	KIDIVVLAFQKASSIVYKKEG	3681	322
B6 (41)	201-221	VLAFQKASSIVYKKEGEQVEF	3566	323
C6 (42)	206-226	KASSIVYKKEGEQVEFSFPLA	14347	324
D6 (43)	211-231	VYKKEGEQVEFSFPLAFTVEK	14740	325
E6 (44)	216-236	GEQVEFSFPLAFTVEKLTGSG	18549	326
F6 (45)	221-241	FSFPLAFTVEKLTGSGELWWQ	9673	327
G6 (46)	226-246	AFTVEKLTGSGELWWQAERAS	3992	328
H6 (47)	231-251	KLTGSGELWWQAERASSSKSW	1878	329
A7 (48)	236-256	GELWWQAERASSSKSWITFDL	2730	330
B7 (49)	241-261	QAERASSSKSWITFDLKNKEV	2588	331
C7 (50)	246-266	SSSKSWITFDLKNKEVSVKRV	1761	332
D7 (51)	251-271	WITFDLKNKEVSVKRVTQDPK	2126	333
E7 (52)	256-276	LKNKEVSVKRVTQDPKLQMGK	2288	334
F7 (53)	261-281	VSVKRVTQDPKLQMGKKLPLH	1848	335
G7 (54)	266-286	VTQDPKLQMGKKLPLHLTLPQ	2075	336
H7 (55)	271-291	KLQMGKKLPLHLTLPQALPQY	1949	337
A8 (56)	276-296	KKLPLHLTLPQALPQYAGSGN	1922	338
B8 (57)	281-301	HLTLPQALPQYAGSGNLTLAL	2394	339
C8 (58)	286-306	QALPQYAGSGNLTLALEAKTG	2364	340
D8 (59)	291-311	YAGSGNLTLALEAKTGKLHQE	1830	341
E8 (60)	296-316	NLTLALEAKTGKLHQEVNLVV	1676	342
F8 (61)	301-321	LEAKTGKLHQEVNLVVMRATQ	1729	343
G8 (62)	306-326	GKLHQEVNLVVMRATQLQKNL	1776	344
H8 (63)	311-331	EVNLVVMRATQLQKNLTCEVW	2183	345
A9 (64)	316-336	VMRATQLQKNLTCEVWGPTSP	2144	346
B9 (65)	321-341	QLQKNLTCEVWGPTSPKLMLS	1856	347
C9 (66)	326-346	LTCEVWGPTSPKLMLSLKLEN	2412	348
D9 (67)	331-351	WGPTSPKLMLSLKLENKEAKV	2414	349
E9 (68)	336-356	PKLMLSLKLENKEAKVSKREK	1656	350
F9 (69)	341-361	SLKLENKEAKVSKREKAVWVL	1663	351
G9 (70)	346-366	NKEAKVSKREKAVWVLNPEAG	1735	352
H9 (71)	351-371	VSKREKAVWVLNPEAGMWQCL	2034	353
A10 (72)	356-376	KAVWVLNPEAGMWQCLLSDSG	3133	354

B10(73)	361-381	LNPEAGMWQCLLSDSGQVLLE	6316	355
C10(74)	366-386	GMWQCLLSDSGQVLLESNIKV	4185	356
D10(75)	371-391	LLSDSGQVLLESNIKVLPTWS	2375	357
E10(76)	376-396	GQVLLESNIKVLPTWSTPVQP	2089	358
F10(77)	381-401	ESNIKVLPTWSTPVQPMALIV	1992	359
G10(78)	386-406	VLPTWSTPVQPMALIVLGGVA	2197	360
H10(79)	391-411	STPVQPMALIVLGGVAGLLLF	2527	361
A11(80)	396-416	PMALIVLGGVAGLLLFIGLGI	3067	362
B11(81)	401-421	VLGGVAGLLLFIGLGIFFCVR	3738	363
C11(82)	406-426	AGLLLFIGLGIFFCVRCRHRR	2099	364
D11(83)	411-431	FIGLGIFFCVRCRHRRRQAER	1900	365
E11(84)	416-436	IFFCVRCRHRRRQAERMSQIK	2085	366
F11(85)	421-441	RCRHRRRQAERMSQIKRLLSE	2075	367
G11(86)	426-446	RRQAERMSQIKRLLSEKKTCQ	1607	368
H11(87)	431-451	RMSQIKRLLSEKKTCQCPHRF	2020	369
A12(88)	436-456	KRLLSEKKTCQCPHRFQKTCS	1674	370
B12(89)	441-458	EKKTCQCPHRFQKTCSPI	2006	371
A1 (0)		empty (control)	2075	

Example 5

This example shows the binding of ^{125}I-HIV-1_{LAI} gp120 to the amino termini of CCR5, CXCR4, and STRL33 as a function of the dependence on position and length. Synthetic peptide arrays of nonapeptides, dodecapeptides, pentadecapeptides and octadecapeptides derived from CCR5 (panel A), CXCR4 (panel B) and STRL33 (panel C) amino terminal domains were prepared and utilized to test the binding of ^{125}I-HIV-1_{LAI} envelope gp120. Ordinal sequence position numbers are given in accordance with the sequence data provided by the Genbank database for CCR5 (accession No. g1457946, gi|1457946), CXCR4 (accession No. g539677, gi|400654, sp|P30991) and STRL33 (accession No. g2209288, gi|2209288). The counts shown are the counts detected in each well minus the background counts (i.e., counts observed in the assay when no polypeptide was bound to the well of the 96-well assay plate).

Panel A CCR5 Initial Sequence #	Peptide Sequence Scanning Windows (In each sequence row 9-, 12-, 15-, 18-mers share the same initial starting point.)	Binding Results for Window Length (counts bound - background (no peptide))				
	xxxxxxxxx 9 xxxxxxxxxxxx 12 xxxxxxxxxxxxxxx 15 xxxxxxxxxxxxxxxxxx 18	9	12	15	18	SEQ ID NO:
1	MDYQVSSPIYDINYYTSE	543	2682	4976	5880	372
2	DYQVSSPIYDINYYTSEP	1552	3089	5401	6363	373
3	YQVSSPIYDINYYTSEPC	2533	5305	5415	6119	374
4	QVSSPIYDINYYTSEPCQ	490	1959	4594	5645	375
5	VSSPIYDINYYTSEPCQK	509	1629	3280	3521	376
6	SSPIYDINYYTSEPCQKI	671	1739	3498	3285	377
7	SPIYDINYYTSEPCQKIN	1503	3463	4575	3234	378
8	PIYDINYYTSEPCQKINV	1186	2285	2682	2036	379
9	IYDINYYTSEPCQKINVK	1359	2702	2516	1261	380
10	YDINYYTSEPCQKINVKQ	4379	5245	3052	1913	381
11	DINYYTSEPCQKINVKQI	1396	1361	1144	712	382
12	INYYTSEPCQKINVKQIA	1384	1190	707	684	383
13	NYYTSEPCQKINVKQIAA	1548	977	760	595	384
14	YYTSEPCQKINVKQIAAR	1029	1052	847	638	385
15	YTSEPCQKINVKQIA	567	507	459		386
16	TSEPCQKINVKQIAA	440	427	509		387
17	SEPCQKINVKQIAAR	434	430	426		388
18	EPCQKINVKQIA	397	432			389
19	PCQKINVKQIAA	386	385			390
20	CQKINVKQIAAR	435	581			391
21	QKINVKQIA	453				392
22	KINVKQIAA	487				393
23	INVKQIAAR	474				394

Panel B CXCR4 Initial Sequence #	Peptide Sequence Scanning Windows (In each sequence row 9-, 12-, 15-, 18-mers share the same initial starting point.)	Binding Results For Window Length (counts bound - background)				
	xxxxxxxxx 9 xxxxxxxxxxxx 12 xxxxxxxxxxxxxxx 15 xxxxxxxxxxxxxxxxxx 18	9	12	15	18	SEQ ID NO:
1	MEGISIYTSDNYTEEMGS	591	334	3275	2079	395
2	EGISIYTSDNYTEEMGSG	[a]	886	7255	1548	396
3	GISIYTSDNYTEEMGSGD	454	2644	3274	1217	397
4	ISIYTSDNYTEEMGSGDY	466	3973	2202	861	398
5	SIYTSDNYTEEMGSGDYD	[a]	288	168	239	399
6	IYTSDNYTEEMGSGDYDS	332	335	195	173	400
7	YTSDNYTEEMGSGDYDSM	181	161	201	103	401
8	TSDNYTEEMGSGDYDSMK	[a]	54	119	38	402
9	SDNYTEEMGSGDYDSMKE	151	149	124	161	403
10	DNYTEEMGSGDYDSMKEP	67	121	57	102	404
11	NYTEEMGSGDYDSMKEPC	[a]	100	30	134	405
12	YTEEMGSGDYDSMKEPCF	68	213	70	103	406
13	TEEMGSGDYDSMKEPCFR	146	67	23	47	407
14	EEMGSGDYDSMKEPCFRE	[a]	61	121	130	408
15	EMGSGDYDSMKEPCFREE	64	36	69	64	409
16	MGSGDYDSMKEPCFREEN	57	68	64	129	410
17	GSGDYDSMKEPCFREENA	[a]	155	172	155	411
18	SGDYDSMKEPCFREENAN	100	118	186	89	412
19	GDYDSMKEPCFREENANF	53	167	198	134	413
20	DYDSMKEPCFREENANFN	[a]	167	146	75	414
21	YDSMKEPCFREENANFNK	171	144	80	89	415
22	DSMKEPCFREENANFNKI	85	144	146	40	416
23	SMKEPCFREENANFN	[a]	119	55		417
24	MKEPCFREENANFNK	188	133	74		418
25	KEPCFREENANFNKI	165	105	93		419
26	EPCFREENANFN	[a]	69			420
27	PCFREENANFNK	104	108			421
28	CFREENANFNKI	103	66			422
29	REENANFNK	58				423

[a] Not done

Panel C STRL33 Initial Sequence #	Peptide Sequence Scanning Windows (In each sequence row 9-, 12-, 15-, 18-mers share the same initial starting point.)	Binding Results For Window Length (counts bound - background)				
	xxxxxxxxx 9 xxxxxxxxxxxx 12 xxxxxxxxxxxxxxx 15 xxxxxxxxxxxxxxxxxx 18	9	12	15	18	SEQ ID NO:
1	MAEHDYHEDYGFSSFNDS	160	625	1239	1386	424
2	AEHDYHEDYGFSSFNDSS	354	697	1095	1014	425
3	EHDYHEDYGFSSFNDSSQ	509	937	2235	1219	426
4	HDYHEDYGFSSFNDSSQE	708	1427	1772	1500	427
5	DYHEDYGFSSFNDSSQEE	851	1554	1240	1191	428
6	YHEDYGFSSFNDSSQEEH	728	1950	1357	985	429
7	HEDYGFSSFNDSSQEEHQ	729	1077	947	537	430
8	EDYGFSSFNDSSQEEHQA	953	817	1152	548	431
9	DYGFSSFNDSSQEEHQAF	701	573	595	440	432
10	YGFSSFNDSSQEEHQAFL	345	745	645	1138	433
11	GFSSFNDSSQEEHQAFLQ	171	480	270	1639	434
12	FSSFNDSSQEEHQAFLQF	249	403	361	3608	435
13	SSFNDSSQEEHQAFLQFS	243	277	902	6038	436
14	SFNDSSQEEHQAFLQFSK	304	303	969	4537	437
15	FNDSSQEEHQAFLQFSKV	246	470	4089	4678	438
16	NDSSQEEHQAFLQFS	180	497	6160		439
17	DSSQEEHQAFLQFSK	147	882	4588		440
18	SSQEEHQAFLQFSKV	287	4455	4732		441
19	SQEEHQAFLQFS	647	7512			442
20	QEEHQAFLQFSK	1109	5672			443
21	EEHQAFLQFSKV	6060	5598			444
22	EHQAFLQFS	7505				445
23	HQAFLQFSK	2761				446
24	QAFLQFSKV	2600				447

Example 6

This example shows ^{125}I-HIV-1_{LAI} gp120 binding to N-terminal peptide variants of CCR5, CXCR4 and STRL33.

Octadecapeptide alanine replacement variants of maximum gp120 binding activity peaks were synthesized and tested for ^{125}I-HIV-1_{LAI} gp120 binding. Each binding value presented is the average of two separate synthesis and

binding experiments. Relative percentage of Control = {[(mean counts/Control counts)] x 100%} ± average deviation. Background counts (no peptide, see Example 7) were subtracted from all values. Data for CCR5 are presented in Panel A; data for CXCR4 are presented in Panel B; and data for STRL33 are presented in Panel C.

Panel A. ^{125}I-HIV-1_{LAI} gp120 binding to N-terminal peptide variants of CCR5

	CCR5 variant peptides (1-18)	Relative % of Control [a]	SEQ ID NO:
Control	MDYQVSSPIYDINYYTSE	100	448
M1A	**A**DYQVSSPIYDINYYTSE	167 ± 4	449
D2A	M**A**YQVSSPIYDINYYTSE	125 ± 8	450
Y3A	MD**A**QVSSPIYDINYYTSE	51 ± 2	451
Q4A	MDY**A**VSSPIYDINYYTSE	104 ± 7	452
V5A	MDYQ**A**SSPIYDINYYTSE	82 ± 3	453
S6A	MDYQV**A**SPIYDINYYTSE	124 ± 3	454
S7A	MDYQVS**A**PIYDINYYTSE	56 ± 2	455
P8A	MDYQVSS**A**IYDINYYTSE	157 ± 2	456
I9A	MDYQVSSP**A**YDINYYTSE	24 ± 7	457
Y10A	MDYQVSSPI**A**DINYYTSE	19 ± 6	458
D11A	MDYQVSSPIY**A**INYYTSE	63 ± 22	459
I12A	MDYQVSSPIYD**A**NYYTSE	14 ± 1	460
N13A	MDYQVSSPIYDI**A**YYTSE	253 ± 19	461
Y14A	MDYQVSSPIYDIN**A**YTSE	15 ± 0.3	462
Y15A	MDYQVSSPIYDINY**A**TSE	21 ± 5	463
T16A	MDYQVSSPIYDINYY**A**SE	78 ± 34	464
S17A	MDYQVSSPIYDINYYT**A**E	64 ± 6	465
E18A	MDYQVSSPIYDINYYTS**A**	4 ± 2	466

[a] The percent binding for the wild-type peptide was defined as 100%.

Panel B ^{125}I-HIV-1$_{LAI}$ gp120 binding to N-terminal peptide variants of CXCR4

	CXCR4 variant peptides (1-18)	Relative % of Control [a]	SEQ ID NO:
Control	MEGISIYTSDNYTEEMGS	100	467
M1A	**A**EGISIYTSDNYTEEMGS	118 ± 18	468
E2A	M**A**GISIYTSDNYTEEMGS	36 ± 0.3	469
G3A	ME**A**ISIYTSDNYTEEMGS	101 ± 3	470
I4A	MEG**A**SIYTSDNYTEEMGS	6 ± 0.3	471
S5A	MEGI**A**IYTSDNYTEEMGS	133 ± 5	472
I6A	MEGIS**A**YTSDNYTEEMGS	2 ± 1	473
Y7A	MEGISI**A**TSDNYTEEMGS	7 ± 0.4	474
T8A	MEGISIY**A**SDNYTEEMGS	97 ± 10	475
S9A	MEGISIYT**A**DNYTEEMGS	70 ± 4	476
D10A	MEGISIYTS**A**NYTEEMGS	71 ± 8	477
N11A	MEGISIYTSD**A**YTEEMGS	38 ± 0.4	478
Y12A	MEGISIYTSDN**A**TEEMGS	28 ± 2	479
T13A	MEGISIYTSDNY**A**EEMGS	70 ± 6	480
E14A	MEGISIYTSDNYT**A**EMGS	72 ± 1	481
E15A	MEGISIYTSDNYTE**A**MGS	56 ± 7	482
M16A	MEGISIYTSDNYTEE**A**GS	88 ± 4	483
G17A	MEGISIYTSDNYTEEM**A**S	68 ± 8	484
S18A	MEGISIYTSDNYTEEMG**A**	79 ± 1	485

[a] The percent binding for the wild-type peptide was defined as 100%.

Panel C ^{125}I-HIV-1_{LAI} gp120 binding to N-terminal peptide variants of STRL33

	STRL33 variant peptides (21-38)	Relative % of Control[a]	SEQ ID NO:
Control	EEHQAFLQFSKVFLPCMY	100	486
E21A	**A**EHQAFLQFSKVFLPCMY	81 ± 2	487
E22A	E**A**HQAFLQFSKVFLPCMY	70 ± 1	488
H23A	EE**A**QAFLQFSKVFLPCMY	99 ± 1	489
Q24A	EEH**A**AFLQFSKVFLPCMY	72 ± 1	490
A25A	EEHQ**A**FLQFSKVFLPCMY	101 ± 1	491
F26A	EEHQA**A**LQFSKVFLPCMY	32 ± 0.1	492
L27A	EEHQAF**A**QFSKVFLPCMY	37 ± 2	493
Q28A	EEHQAFL**A**FSKVFLPCMY	44 ± 0.4	494
F29A	EEHQAFLQ**A**SKVFLPCMY	20 ± 1	495
S30A	EEHQAFLQF**A**KVFLPCMY	92 ± 2	496
K31A	EEHQAFLQFS**A**VFLPCMY	162 ± 2	497
V32A	EEHQAFLQFSK**A**FLPCMY	51 ± 3	498
F33A	EEHQAFLQFSKV**A**LPCMY	45 ± 2	499
L34A	EEHQAFLQFSKVF**A**PCMY	76 ± 1	500
P35A	EEHQAFLQFSKVFL**A**CMY	82 ± 3	501
C36A	EEHQAFLQFSKVFLP**A**MY	53 ± 5	502
M37A	EEHQAFLQFSKVFLPC**A**Y	112 ± 4	503
Y38A	EEHQAFLQFSKVFLPCM**A**	83 ± 2	504

[a] The percent binding for the wild-type peptide was defined as 100%.

Example 7

This example demonstrates that the binding of HIV-1 gp120 envelope protein to the polypeptides of the present invention and to the chemokine receptors from which the present inventive polypeptides were originally derived or inspired is conserved across the various species of HIV-1. This example also demonstrates that a step subsequent to initial binding of gp120 to CCR5, CXCR4, STRL33, and CD4 is the most likely source of the phenomenon of host-range selectivity. Additionally, this example demonstrates that the underlying method is accurate in that receptor variants that are predicted to have an altered affinity for binding with gp120, do in fact have a statistically similar alteration in affinity where comparable changes in the receptors have been identified in other work and the

affinity for binding of gp120/effect on infectivity has been measured.

This example examines the effect of particular mutations of CCR5 that were studied in the work underlying the present invention and that were also studied by other artisans in the field.

The following table identifies a mutation in the first column. The first letter designates the wild-type amino acid present at the position indicated by the number, and the letter A which terminates all entries in the first column indicates that the amino acid residue present in that position in the mutant polypeptide is alaninyl. For example, the first data row (i.e., the second row of the table) contains the entry Y3A in the first column, which indicates that the tyrosine residue at position 3 of the wild-type CCR5 is substituted by an alanine residue.

The second column provides the percentage of binding exhibited by a mutant polypeptide compared to a wild-type polypeptide, when the methods used to elucidate the present invention are used in conjunction with radiolabeled HIV-1_{LAI} gp120 envelope protein. The third through seventh columns provide similar data that have been extracted from the work of others in the field using a strain of HIV-1 virus indicated at the top of each column. For example, row 2 of the following table indicates that when the mutation Y3A is effected in the human CCR5 chemokine receptor, then the resulting CCR5 polypeptide has 51.4% of the ability to bind HIV-1_{LAI} gp120 envelope protein in comparison to an equivalent wild-type peptide. Similarly, HIV-1_{ADA} binds to the mutant polypeptide with 79% of the affinity of a non-mutated CCR5 chemokine receptor.

	gp120	YU2	ADA	JF-RL	89.6	DH123
Y3A	51.4	n/a	79	82	n/a	42
Q4A	104	85	132	111	67	105
Y10A	19.2	2	50	26	10	3
D11A	62.8	2	27	22	6	3
Y14A	14.6	12	47	25	6	0
Y15A	21	30	3	3	1	0
E18A	4.1	45	12	12	3	10

Statistical analysis of these data indicates that the similarity between the binding affinity of each mutant peptide for gp120 elucidated in this study is not more than about 25% likely to be causally unrelated to the effects observed for YU2, and not more than about 4% likely to be causally unrelated to the effects observed for each of the other viruses listed in the table above.

Additionally, the affinity measurements generated by the underlying technique has been demonstrated to be accurate by (repetitively) showing that antibodies that specifically bind to radiolabeled gp120 are capable of preventing the binding of gp120 to polypeptides that have shown high affinity for binding with gp120 in the experiments upon which the present invention is predicated. Thus, this example shows that the binding with chemokine receptors HIV-1 can be inhibited by the present inventive polypeptides, irrespective of the strain of HIV-1 from which the gp120 protein is obtained.

Example 8

This example provides a characterization of the critical amino acids in the amino-terminal segments of CCR5, CXCR4, and STRL33 that are essential for the ability of these polypeptides to bind with gp120.

In this example, the effect on binding that occurs to due successive replacement of each amino acid with alanine is indicated, wherein a (+) signifies a decrease in binding affinity and a (>) signifies an enhancement in binding

affinity. As is clear from inspection, the sequences are shown with that amino-terminus at top and the carboxyl-terminus at bottom.

CCR5 (1-18)	CXCR4 (1-18)	STRL33 (21-38)
M>	M	E
D	E+	E
Y++	G	H
Q	I+++++	Q
V	S>	A
S	I++++++	F+++
S+	Y+++++	L++
P>	T	Q+
I+++	S+	F+++
Y+++	D+	S
D+	N++	K>
I++++	Y++	V+
N>	T	F+
Y++++	E	L
Y+++	E++	P
T	M	C+
S+	G	M
E+++++	S	Y

Example 9

This example employs the same technique as Example 4 and provides information similar to that available from Example 4.

The data below compares the ability of synthetic fragments of CD4 to bind to labeled gp120. 9-mer, 12-mer, 15-mer, 18-mer, and 21-mers were selected based on the data from Examples 4. The relative binding affinities of each group of polypeptides can be determined by inspection of the number of counts of radiolabeled gp120 that were retained by each N-mer. Data supporting these conclusions are provided by Examples 10 and 11.

Peptide starting position #	Active Peptides	gp120 bound (counts)	SEQ ID NO:	Peptide starting position #	Active Peptides	Gp120 Bound (counts)	SEQ ID NO:
	ACTIVE 9-MERS				ACTIVE 12-MERS		
105	DTYICEVED	1043	505	101	IEDSDTYICEVE	1107	530
115	KEEVQLLVF	1273	506	112	EDQKEEVQLLVF	1379	531
116	EEVQLLVFG	3170	507	113	DQKEEVQLLVFG	1624	532
117	EVQLLVFGL	2146	508	114	QKEEVQLLVFGL	1785	533
				115	KEEVQLLVFGLT	1774	534
				116	EEVQLLVFGLTA	3261	535
				117	EVQLLVFGLTAN	1838	536
				133	LLQGQSLTLTLE	1320	537
217	EQVEFSFPL	1032	509	215	EGEQVEFSFPLA	1456	538
218	QVEFSFPLA	1205	510	216	GEQVEFSFPLAF	1729	539
219	VEFSFPLAF	1064	511	217	EQVEFSFPLAFT	1556	540
				218	QVEFSFPLAFTV	1636	541
	ACTIVE 15-MERS				ACTIVE 18-MERS		
109	CEVEDQKEEVQLLVF	1729	512	105	DTYICEVEDQKEEVQLLV	1648	542
110	EVEDQKEEVQLLVFG	2805	513	106	TYICEVEDQKEEVQLLVF	3794	543
111	VEDQKEEVQLLVFGL	3816	514	107	YICEVEDQKEEVQLLVFG	4611	544
112	EDQKEEVQLLVFGLT	3633	515	108	ICEVEDQKEEVQLLVFGL	3898	545
113	DQKEEVQLLVFGLTA	3905	516	109	CEVEDQKEEVQLLVFGLT	3797	546
114	QKEEVQLLVFGLTAN	3770	517	110	EVEDQKEEVQLLVFGLTA	3647	547
115	KEEVQLLVFGLTANS	3485	518	111	VEDQKEEVQLLVFGLTAN	3913	548
116	EEVQLLVFGLTANSD	6423	519	112	EDQKEEVQLLVFGLTANS	3416	549
117	EVQLLVFGLTANSDT	2689	520	113	DQKEEVQLLVFGLTANSD	3317	550
				114	QKEEVQLLVFGLTANSDT	3671	551
130	DTHLLQGQSLTLTLE	1622	521	127	ANSDTHLLQGQSLTLTLE	1540	552
131	THLLQGQSLTLTLES	1874	522	128	NSDTHLLQGQSLTLTLES	1726	553
132	HLLQGQSLTLTLESP	1277	523	129	SDTHLLQGQSLTLTLESP	1260	554
213	KKEGEQVEFSFPLAF	1921	524	210	IVYKKEGEQVEFSFPLAF	5382	555
214	KEGEQVEFSFPLAFT	3253	525	211	VYKKEGEQVEFSFPLAFT	4307	556
215	EGEQVEFSFPLAFTV	3270	526	212	YKKEGEQVEFSFPLAFTV	4839	557
216	GEQVEFSFPLAFTVE	4656	527	213	KKEGEQVEFSFPLAFTVE	4683	558
217	EQVEFSFPLAFTVEK	4135	528	214	KEGEQVEFSFPLAFTVEK	3117	559
218	QVEFSFPLAFTVEKL	2047	529	215	EGEQVEFSFPLAFTVEKL	2164	560
				216	GEQVEFSFPLAFTVEKLT	1643	561
	ACTIVE 21-MERS						
90	GNFPLIIKNLKIEDSDTYICE	5248	562				
91	NFPLIIKNLKIEDSDTYICEV	7803	563				
92	FPLIIKNLKIEDSDTYICEVE	13919	564				

93	PLIIKNLKIEDSDTYICEVED	20145	565
94	LIIKNLKIEDSDTYICEVEDQ	17108	566
95	IIKNLKIEDSDTYICEVEDQK	11892	567
96	IKNLKIEDSDTYICEVEDQKE	15073	568
97	KNLKIEDSDTYICEVEDQKEE	8789	569
99	LKIEDSDTYICEVEDQKEEVQ	5519	570
100	KIEDSDTYICEVEDQKEEVQL	6325	571
101	IEDSDTYICEVEDQKEEVQLL	12064	572
102	EDSDTYICEVEDQKEEVQLLV	4933	573
103	DSDTYICEVEDQKEEVQLLVF	30277	574
104	SDTYICEVEDQKEEVQLLVFG	30319	575
105	DTYICEVEDQKEEVQLLVFGL	25424	576
106	TYICEVEDQKEEVQLLVFGLT	20191	577
107	YICEVEDQKEEVQLLVFGLTA	22884	578
108	ICEVEDQKEEVQLLVFGLTAN	7276	579
109	CEVEDQKEEVQLLVFGLTANS	3517	580
123	FGLTANSDTHLLQGQSLTLTL	11529	581
124	GLTANSDTHLLQGQSLTLTLE	14065	582
125	LTANSDTHLLQGQSLTLTLES	17113	583
126	TANSDTHLLQGQSLTLTLESP	23595	584
204	FQKASSIVYKKEGEQVEFSFP	9382	585
205	QKASSIVYKKEGEQVEFSFPL	24959	586
206	KASSIVYKKEGEQVEFSFPLA	30873	587
207	ASSIVYKKEGEQVEFSFPLAF	25146	588
208	SSIVYKKEGEQVEFSFPLAFT	28068	589
209	SIVYKKEGEQVEFSFPLAFTV	8165	590
210	IVYKKEGEQVEFSFPLAFTVE	15620	591
221	FSFPLAFTVEKLTGSGELWWQ	4163	592
222	SFPLAFTVEKLTGSGELWWQA	2284	593
223	FPLAFTVEKLTGSGELWWQAE	6276	594
224	PLAFTVEKLTGSGELWWQAER	2647	595
225	LAFTVEKLTGSGELWWQAERA	3577	596

Example 10

This example provides data which enables those skilled in the art to arrive at the conclusions indicated in Examples 9 and 12. In this example, the counts of radiolabeled gp-120 retained by each peptide indicated in the left hand column are given in the right hand column. The first panel (panel A) provides data for 21-mers of CD4.

Panel A

PEPTIDE	COUNTS	SEQ ID NO:
LWDQGNFPLIIKNLKIEDSDT	731	597
WDQGNFPLIIKNLKIEDSDTY	889	598
DQGNFPLIIKNLKIEDSDTYI	1138	599
QGNFPLIIKNLKIEDSDTYIC	2242	600
GNFPLIIKNLKIEDSDTYICE	5248	601
NFPLIIKNLKIEDSDTYICEV	7803	602
FPLIIKNLKIEDSDTYICEVE	13919	603
PLIIKNLKIEDSDTYICEVED	20145	604
LIIKNLKIEDSDTYICEVEDQ	17108	605
IIKNLKIEDSDTYICEVEDQK	11892	606
IKNLKIEDSDTYICEVEDQKE	15073	607
KNLKIEDSDTYICEVEDQKEE	8789	608
NLKIEDSDTYICEVEDQKEEV	2016	609
LKIEDSDTYICEVEDQKEEVQ	5519	610
KIEDSDTYICEVEDQKEEVQL	6325	611
IEDSDTYICEVEDQKEEVQLL	12064	612
EDSDTYICEVEDQKEEVQLLV	4933	613
DSDTYICEVEDQKEEVQLLVF	30277	614
SDTYICEVEDQKEEVQLLVFG	30319	615
DTYICEVEDQKEEVQLLVFGL	25424	616
TYICEVEDQKEEVQLLVFGLT	20191	617
YICEVEDQKEEVQLLVFGLTA	22884	618
ICEVEDQKEEVQLLVFGLTAN	7276	619
CEVEDQKEEVQLLVFGLTANS	3517	620
EVEDQKEEVQLLVFGLTANSD	1687	621
VEDQKEEVQLLVFGLTANSDT	646	622
EDQKEEVQLLVFGLTANSDTH	562	623
DQKEEVQLLVFGLTANSDTHL	599	624
QKEEVQLLVFGLTANSDTHLL	573	625
KEEVQLLVFGLTANSDTHLLQ	682	626
EEVQLLVFGLTANSDTHLLQG	690	627
EVQLLVFGLTANSDTHLLQGQ	589	628
VQLLVFGLTANSDTHLLQGQS	1099	629
QLLVFGLTANSDTHLLQGQSL	2057	630
LLVFGLTANSDTHLLQGQSLT	860	631
LVFGLTANSDTHLLQGQSLTL	4677	632
VFGLTANSDTHLLQGQSLTLT	2762	633
FGLTANSDTHLLQGQSLTLTL	11529	634
GLTANSDTHLLQGQSLTLTLE	14065	635
LTANSDTHLLQGQSLTLTLES	17113	636
TANSDTHLLQGQSLTLTLESP	23595	637
Empty (Control)	515	
TWTCTVLQNQKKVEFKIDIVV	1430	638
WTCTVLQNQKKVEFKIDIVVL	1616	639
TCTVLQNQKKVEFKIDIVVLA	1092	640

CTVLQNQKKVEFKIDIVVLAF	2909	641
TVLQNQKKVEFKIDIVVLAFQ	3273	642
VLQNQKKVEFKIDIVVLAFQK	1323	643
LQNQKKVEFKIDIVVLAFQKA	1256	644
QNQKKVEFKIDIVVLAFQKAS	1808	645
NQKKVEFKIDIVVLAFQKASS	1507	646
QKKVEFKIDIVVLAFQKASSI	759	647
KKVEFKIDIVVLAFQKASSIV	782	648
KVEFKIDIVVLAFQKASSIVY	635	649
VEFKIDIVVLAFQKASSIVYK	725	650
EFKIDIVVLAFQKASSIVYKK	649	651
FKIDIVVLAFQKASSIVYKKE	593	652
KIDIVVLAFQKASSIVYKKEG	1394	653
IDIVVLAFQKASSIVYKKEGE	962	654
DIVVLAFQKASSIVYKKEGEQ	788	655
IVVLAFQKASSIVYKKEGEQV	646	656
VVLAFQKASSIVYKKEGEQVE	772	657
VLAFQKASSIVYKKEGEQVEF	1793	658
LAFQKASSIVYKKEGEQVEFS	1410	659
AFQKASSIVYKKEGEQVEFSF	3775	660
FQKASSIVYKKEGEQVEFSFP	9382	661
QKASSIVYKKEGEQVEFSFPL	24959	662
KASSIVYKKEGEQVEFSFPLA	30873	663
ASSIVYKKEGEQVEFSFPLAF	25146	664
SSIVYKKEGEQVEFSFPLAFT	28068	665
SIVYKKEGEQVEFSFPLAFTV	8165	666
IVYKKEGEQVEFSFPLAFTVE	15620	667
VYKKEGEQVEFSFPLAFTVEK	2429	668
YKKEGEQVEFSFPLAFTVEKL	735	669
KKEGEQVEFSFPLAFTVEKLT	1847	670
KEGEQVEFSFPLAFTVEKLTG	972	671
EGEQVEFSFPLAFTVEKLTGS	739	672
GEQVEFSFPLAFTVEKLTGSG	652	673
EQVEFSFPLAFTVEKLTGSGE	765	674
QVEFSFPLAFTVEKLTGSGEL	741	675
VEFSFPLAFTVEKLTGSGELW	633	676
EFSFPLAFTVEKLTGSGELWW	681	677
FSFPLAFTVEKLTGSGELWWQ	4163	678
SFPLAFTVEKLTGSGELWWQA	2284	679
FPLAFTVEKLTGSGELWWQAE	6276	680
PLAFTVEKLTGSGELWWQAER	2647	681
LAFTVEKLTGSGELWWQAERA	3577	682
AFTVEKLTGSGELWWQAERAS	1739	683
Empty (control)	617	

These second and third panels (panels B and C) provide data for 18-mers of a small region of CD4.

Panel B

PEPTIDE	COUNTS	SEQ ID NO:
LWDQGNFPLIIKNLK	502	684
WDQGNFPLIIKNLKI	534	685
DQGNFPLIIKNLKIE	635	686
QGNFPLIIKNLKIED	509	687
GNFPLIIKNLKIEDS	624	688
NFPLIIKNLKIEDSD	654	689
FPLIIKNLKIEDSDT	539	690
PLIIKNLKIEDSDTY	661	691
LIIKNLKIEDSDTYI	542	692
IIKNLKIEDSDTYIC	664	693
IKNLKIEDSDTYICE	568	694
KNLKIEDSDTYICEV	562	695
NLKIEDSDTYICEVE	1160	696
LKIEDSDTYICEVED	846	697
KIEDSDTYICEVEDQ	1088	698
IEDSDTYICEVEDQK	1143	699
EDSDTYICEVEDQKE	815	700
DSDTYICEVEDQKEE	973	701
SDTYICEVEDQKEEV	993	702
DTYICEVEDQKEEVQ	1071	703
TYICEVEDQKEEVQL	956	704
YICEVEDQKEEVQLL	1064	705
ICEVEDQKEEVQLLV	1084	706
CEVEDQKEEVQLLVF	1729	707
EVEDQKEEVQLLVFG	2805	708
VEDQKEEVQLLVFGL	3816	709
EDQKEEVQLLVFGLT	3633	710
DQKEEVQLLVFGLTA	3905	711
QKEEVQLLVFGLTAN	3770	712
KEEVQLLVFGLTANS	3485	713
EEVQLLVFGLTANSD	6423	714
EVQLLVFGLTANSDT	2689	715
VQLLVFGLTANSDTH	1006	716
QLLVFGLTANSDTHL	865	717
LLVFGLTANSDTHLL	599	718
LVFGLTANSDTHLLQ	609	719
VFGLTANSDTHLLQG	532	720
FGLTANSDTHLLQGQ	625	721
GLTANSDTHLLQGQS	532	722
LTANSDTHLLQGQSL	634	723
TANSDTHLLQGQSLT	513	724
ANSDTHLLQGQSLTL	542	725

NSDTHLLQGQSLTLT	631	726
SDTHLLQGQSLTLTL	747	727
DTHLLQGQSLTLTLE	1622	728
THLLQGQSLTLTLES	1874	729
HLLQGQSLTLTLESP	1277	730
LWDQGNFPLIIKNLKIED	582	731
WDQGNFPLIIKNLKIEDS	626	732
DQGNFPLIIKNLKIEDSD	598	733
QGNFPLIIKNLKIEDSDT	564	734
GNFPLIIKNLKIEDSDTY	557	735
NFPLIIKNLKIEDSDTYI	627	736
FPLIIKNLKIEDSDTYIC	509	737
PLIIKNLKIEDSDTYICE	624	738
LIIKNLKIEDSDTYICEV	634	739
IIKNLKIEDSDTYICEVE	751	740
IKNLKIEDSDTYICEVED	699	741
KNLKIEDSDTYICEVEDQ	708	742
NLKIEDSDTYICEVEDQK	863	743
LKIEDSDTYICEVEDQKE	872	744
KIEDSDTYICEVEDQKEE	858	745
IEDSDTYICEVEDQKEEV	1230	746
EDSDTYICEVEDQKEEVQ	788	747
DSDTYICEVEDQKEEVQL	961	748
SDTYICEVEDQKEEVQLL	870	749
DTYICEVEDQKEEVQLLV	1648	750
TYICEVEDQKEEVQLLVF	3794	751
YICEVEDQKEEVQLLVFG	4611	752
ICEVEDQKEEVQLLVFGL	3898	753
CEVEDQKEEVQLLVFGLT	3797	754
EVEDQKEEVQLLVFGLTA	3647	755
VEDQKEEVQLLVFGLTAN	3913	756
EDQKEEVQLLVFGLTANS	3416	757
DQKEEVQLLVFGLTANSD	3317	758
QKEEVQLLVFGLTANSDT	3671	759
KEEVQLLVFGLTANSDTH	1271	760
EEVQLLVFGLTANSDTHL	783	761
EVQLLVFGLTANSDTHLL	667	762
VQLLVFGLTANSDTHLLQ	673	763
QLLVFGLTANSDTHLLQG	574	764
LLVFGLTANSDTHLLQGQ	568	765
LVFGLTANSDTHLLQGQS	564	766
VFGLTANSDTHLLQGQSL	531	767
FGLTANSDTHLLQGQSLT	591	768
GLTANSDTHLLQGQSLTL	572	769
LTANSDTHLLQGQSLTLT	528	770
TANSDTHLLQGQSLTLTL	891	771
ANSDTHLLQGQSLTLTLE	1540	772
NSDTHLLQGQSLTLTLES	1726	773

SDTHLLQGQSLTLTLESP	1260	774
Empty (control)	575	

Panel C

PEPTIDE	COUNTS	SEQ ID NO:
WTCTVLQNQKKVEFK	566	775
TCTVLQNQKKVEFKI	510	776
CTVLQNQKKVEFKID	608	777
TVLQNQKKVEFKIDI	587	778
VLQNQKKVEFKIDIV	605	779
LQNQKKVEFKIDIVV	644	780
QNQKKVEFKIDIVVL	636	781
NQKKVEFKIDIVVLA	860	782
QKKVEFKIDIVVLAF	1333	783
KKVEFKIDIVVLAFQ	951	784
KVEFKIDIVVLAFQK	1051	785
VEFKIDIVVLAFQKA	1005	786
EFKIDIVVLAFQKAS	1188	787
FKIDIVVLAFQKASS	1001	788
KIDIVVLAFQKASSI	956	789
IDIVVLAFQKASSIV	865	790
DIVVLAFQKASSIVY	776	791
IVVLAFQKASSIVYK	783	792
VVLAFQKASSIVYKK	577	793
VLAFQKASSIVYKKE	634	794
LAFQKASSIVYKKEG	593	795
AFQKASSIVYKKEGE	544	796
FQKASSIVYKKEGEQ	637	797
QKASSIVYKKEGEQV	519	798
KASSIVYKKEGEQVE	563	799
ASSIVYKKEGEQVEF	589	800
SSIVYKKEGEQVEFS	558	801
SIVYKKEGEQVEFSF	651	802
IVYKKEGEQVEFSFP	615	803
VYKKEGEQVEFSFPL	714	804
YKKEGEQVEFSFPLA	687	805
KKEGEQVEFSFPLAF	1921	806
KEGEQVEFSFPLAFT	3253	807
EGEQVEFSFPLAFTV	3270	808
GEQVEFSFPLAFTVE	4656	809
EQVEFSFPLAFTVEK	4135	810
QVEFSFPLAFTVEKL	2047	811
VEFSFPLAFTVEKLT	899	812
EFSFPLAFTVEKLTG	920	813
FSFPLAFTVEKLTGS	672	814
SFPLAFTVEKLTGSG	565	815

FPLAFTVEKLTGSGE	556	816
PLAFTVEKLTGSGEL	612	817
LAFTVEKLTGSGELW	579	818
AFTVEKLTGSGELWW	586	819
FTVEKLTGSGELWWQ	625	820
TVEKLTGSGELWWQA	550	821
VEKLTGSGELWWQAE	735	822
EKLTGSGELWWQAER	683	823
WTCTVLQNQKKVEFKIDI	588	824
TCTVLQNQKKVEFKIDIV	571	825
CTVLQNQKKVEFKIDIVV	553	826
TVLQNQKKVEFKIDIVVL	655	827
VLQNQKKVEFKIDIVVLA	724	828
LQNQKKVEFKIDIVVLAF	938	829
QNQKKVEFKIDIVVLAFQ	917	830
NQKKVEFKIDIVVLAFQK	889	831
QKKVEFKIDIVVLAFQKA	1013	832
KKVEFKIDIVVLAFQKAS	912	833
KVEFKIDIVVLAFQKASS	1011	834
VEFKIDIVVLAFQKASSI	819	835
EFKIDIVVLAFQKASSIV	799	836
FKIDIVVLAFQKASSIVY	843	837
KIDIVVLAFQKASSIVYK	779	838
IDIVVLAFQKASSIVYKK	711	839
DIVVLAFQKASSIVYKKE	660	840
IVVLAFQKASSIVYKKEG	531	841
VVLAFQKASSIVYKKEGE	560	842
VLAFQKASSIVYKKEGEQ	549	843
LAFQKASSIVYKKEGEQV	665	844
AFQKASSIVYKKEGEQVE	514	845
FQKASSIVYKKEGEQVEF	528	846
QKASSIVYKKEGEQVEFS	602	847
KASSIVYKKEGEQVEFSF	536	848
ASSIVYKKEGEQVEFSFP	701	849
SSIVYKKEGEQVEFSFPL	756	850
SIVYKKEGEQVEFSFPLA	771	851
IVYKKEGEQVEFSFPLAF	5382	852
VYKKEGEQVEFSFPLAFT	4307	853
YKKEGEQVEFSFPLAFTV	4839	854
KKEGEQVEFSFPLAFTVE	4683	855
KEGEQVEFSFPLAFTVEK	3117	856
EGEQVEFSFPLAFTVEKL	2164	857
GEQVEFSFPLAFTVEKLT	1643	858
EQVEFSFPLAFTVEKLTG	798	859
QVEFSFPLAFTVEKLTGS	736	860
VEFSFPLAFTVEKLTGSG	533	861
EFSFPLAFTVEKLTGSGE	668	862
FSFPLAFTVEKLTGSGEL	613	863

SFPLAFTVEKLTGSGELW	656	864
FPLAFTVEKLTGSGELWW	586	865
PLAFTVEKLTGSGELWWQ	650	866
LAFTVEKLTGSGELWWQA	866	867
AFTVEKLTGSGELWWQAE	788	868
FTVEKLTGSGELWWQAER	1143	869
Empty (control)	556	

The fourth and fifth panels (Panels D and E) provide data for select 9-mers and 12-mers of CD4.

Panel D

PEPTIDE	COUNTS	SEQ ID NO:
DQGNFPLII	662	870
QGNFPLIIK	508	871
GNFPLIIKN	600	872
NFPLIIKNL	561	873
FPLIIKNLK	601	874
PLIIKNLKI	697	875
LIIKNLKIE	515	876
IIKNLKIED	658	877
IKNLKIEDS	557	878
KNLKIEDSD	612	879
NLKIEDSDT	512	880
LKIEDSDTY	492	881
KIEDSDTYI	603	882
IEDSDTYIC	567	883
EDSDTYICE	650	884
DSDTYICEV	712	885
SDTYICEVE	819	886
DTYICEVED	1043	887
TYICEVEDQ	805	888
YICEVEDQK	728	889
ICEVEDQKE	596	890
CEVEDQKEE	555	891
EVEDQKEEV	587	892
VEDQKEEVQ	521	893
EDQKEEVQL	564	894
DQKEEVQLL	589	895
QKEEVQLLV	636	896
KEEVQLLVF	1273	897
EEVQLLVFG	3170	898
EVQLLVFGL	2146	899
VQLLVFGLT	815	900
QLLVFGLTA	822	901
LLVFGLTAN	576	902

LVFGLTANS	522	903
VFGLTANSD	549	904
FGLTANSDT	563	905
GLTANSDTH	481	906
LTANSDTHL	596	907
TANSDTHLL	554	908
ANSDTHLLQ	642	909
NSDTHLLQG	561	910
SDTHLLQGQ	526	911
DTHLLQGQS	578	912
THLLQGQSL	512	913
HLLQGQSLT	564	914
LLQGQSLTL	568	915
LQGQSLTLT	501	916
QGQSLTLTL	594	917
GQSLTLTLE	777	918
DQGNFPLIIKNL	604	919
QGNFPLIIKNLK	533	920
GNFPLIIKNLKI	547	921
NFPLIIKNLKIE	647	922
FPLIIKNLKIED	511	923
PLIIKNLKIEDS	565	924
LIIKNLKIEDSD	619	925
IIKNLKIEDSDT	511	926
IKNLKIEDSDTY	574	927
KNLKIEDSDTYI	523	928
NLKIEDSDTYIC	639	929
LKIEDSDTYICE	635	930
KIEDSDTYICEV	601	931
IEDSDTYICEVE	1107	932
EDSDTYICEVED	956	933
DSDTYICEVEDQ	937	934
SDTYICEVEDQK	846	935
DTYICEVEDQKE	720	936
TYICEVEDQKEE	818	937
YICEVEDQKEEV	734	938
ICEVEDQKEEVQ	585	939
CEVEDQKEEVQL	561	940
EVEDQKEEVQLL	508	941
VEDQKEEVQLLV	657	942
EDQKEEVQLLVF	1379	943
DQKEEVQLLVFG	1624	944
QKEEVQLLVFGL	1785	945
KEEVQLLVFGLT	1774	946
EEVQLLVFGLTA	3261	947
EVQLLVFGLTAN	1838	948
VQLLVFGLTANS	747	949
QLLVFGLTANSD	721	950

LLVFGLTANSDT	533	951
LVFGLTANSDTH	586	952
VFGLTANSDTHL	548	953
FGLTANSDTHLL	571	954
GLTANSDTHLLQ	574	955
LTANSDTHLLQG	534	956
TANSDTHLLQGQ	549	957
ANSDTHLLQGQS	559	958
NSDTHLLQGQSL	585	959
SDTHLLQGQSLT	540	960
DTHLLQGQSLTL	527	961
THLLQGQSLTLT	646	962
HLLQGQSLTLTL	701	963
LLQGQSLTLTLE	1320	964
Empty (control)	581	

Panel E

PEPTIDE	COUNTS	SEQ ID NO:
TVLQNQKKV	534	965
VLQNQKKVE	556	966
LQNQKKVEF	565	967
QNQKKVEFK	537	968
NQKKVEFKI	597	969
QKKVEFKID	575	970
KKVEFKIDI	501	971
KVEFKIDIV	555	972
VEFKIDIVV	548	973
EFKIDIVVL	665	974
FKIDIVVLA	568	975
KIDIVVLAF	665	976
IDIVVLAFQ	691	977
DIVVLAFQK	686	978
IVVLAFQKA	602	979
VVLAFQKAS	600	980
VLAFQKASS	466	981
LAFQKASSI	592	982
AFQKASSIV	595	983
FQKASSIVY	568	984
QKASSIVYK	494	985
KASSIVYKK	498	986
ASSIVYKKE	600	987
SSIVYKKEG	515	988
SIVYKKEGE	566	989
IVYKKEGEQ	534	990
VYKKEGEQV	490	991
YKKEGEQVE	518	992

KKEGEQVEF	546	993
KEGEQVEFS	595	994
EGEQVEFSF	735	995
GEQVEFSFP	697	996
EQVEFSFPL	1032	997
QVEFSFPLA	1205	998
VEFSFPLAF	1064	999
EFSFPLAFT	658	1000
FSFPLAFTV	472	1001
SFPLAFTVE	619	1002
FPLAFTVEK	569	1003
PLAFTVEKL	597	1004
LAFTVEKLT	501	1005
AFTVEKLTG	517	1006
FTVEKLTGS	574	1007
TVEKLTGSG	487	1008
VEKLTGSGE	585	1009
EKLTGSGEL	541	1010
KLTGSGELW	491	1011
LTGSGELWW	550	1012
TGSGELWWQ	507	1013
TVLQNQKKVEFK	563	1014
VLQNQKKVEFKI	503	1015
LQNQKKVEFKID	508	1016
QNQKKVEFKIDI	559	1017
NQKKVEFKIDIV	532	1018
QKKVEFKIDIVV	595	1019
KKVEFKIDIVVL	597	1020
KVEFKIDIVVLA	560	1021
VEFKIDIVVLAF	681	1022
EFKIDIVVLAFQ	659	1023
FKIDIVVLAFQK	736	1024
KIDIVVLAFQKA	689	1025
IDIVVLAFQKAS	630	1026
DIVVLAFQKASS	746	1027
IVVLAFQKASSI	548	1028
VVLAFQKASSIV	567	1029
VLAFQKASSIVY	548	1030
LAFQKASSIVYK	465	1031
AFQKASSIVYKK	597	1032
FQKASSIVYKKE	577	1033
QKASSIVYKKEG	596	1034
KASSIVYKKEGE	559	1035
ASSIVYKKEGEQ	523	1036
SSIVYKKEGEQV	615	1037
SIVYKKEGEQVE	543	1038
IVYKKEGEQVEF	533	1039
VYKKEGEQVEFS	584	1040

YKKEGEQVEFSF	548	1041
KKEGEQVEFSFP	598	1042
KEGEQVEFSFPL	710	1043
EGEQVEFSFPLA	1456	1044
GEQVEFSFPLAF	1729	1045
EQVEFSFPLAFT	1556	1046
QVEFSFPLAFTV	1636	1047
VEFSFPLAFTVE	518	1048
EFSFPLAFTVEK	585	1049
FSFPLAFTVEKL	573	1050
SFPLAFTVEKLT	528	1051
FPLAFTVEKLTG	622	1052
PLAFTVEKLTGS	528	1053
LAFTVEKLTGSG	608	1054
AFTVEKLTGSGE	511	1055
FTVEKLTGSGEL	530	1056
TVEKLTGSGELW	573	1057
VEKLTGSGELWW	477	1058
EKLTGSGELWWQ	543	1059
Empty (control)	571	

Panels F and G provide data on sequential alanine replacements for selected CD4 polypeptides.

Panel F

PEPTIDE	COUNTS	SEQ ID NO:
ZZZZZZDTYICEVED	5844	1060
ZZZZZZATYICEVED	5921	1061
ZZZZZZDAYICEVED	6362	1062
ZZZZZZDTAICEVED	1301	1063
ZZZZZZDTYACEVED	2583	1064
ZZZZZZDTYIAEVED	4483	1065
ZZZZZZDTYICAVED	3154	1066
ZZZZZZDTYICEAED	3432	1067
ZZZZZZDTYICEVAD	3595	1068
ZZZZZZDTYICEVEA	5942	1069
ZZZZZZDTYICEVED	4973	1070
ZZZZZZDTYICEVED	4775	1070
ZZZZZZATYICEVED	4962	1071
ZZZZZZDAYICEVED	4163	1072
ZZZZZZDTAICEVED	1384	1073
ZZZZZZDTYACEVED	3085	1074
ZZZZZZDTYIAEVED	5128	1075
ZZZZZZDTYICAVED	2587	1076
ZZZZZZDTYICEAED	2499	1077

ZZZZZZDTYICEVAD	2706	1078
ZZZZZZDTYICEVEA	6345	1079
ZZZZZZDTYICEVED	5564	1080
EEVQLLVFGLTANSD	18582	1081
AEVQLLVFGLTANSD	16220	1082
EAVQLLVFGLTANSD	14220	1083
EEAQLLVFGLTANSD	18124	1084
EEVALLVFGLTANSD	10890	1085
EEVQALVFGLTANSD	11258	1086
EEVQLAVFGLTANSD	11954	1087
EEVQLLAFGLTANSD	13317	1088
EEVQLLVAGLTANSD	9573	1089
EEVQLLVFALTANSD	19348	1090
EEVQLLVFGATANSD	10408	1091
EEVQLLVFGLAANSD	19973	1092
EEVQLLVFGLTTNSD	20100	1093
EEVQLLVFGLTAASD	19390	1094
EEVQLLVFGLTANAD	17684	1095
EEVQLLVFGLTANSA	18227	1096
EEVQLLVFGLTANSD	19738	1097
EEVQLLVFGLTANSD	21338	1098
AEVQLLVFGLTANSD	14590	1099
EAVQLLVFGLTANSD	13213	1100
EEAQLLVFGLTANSD	16296	1101
EEVALLVFGLTANSD	13415	1102
EEVQALVFGLTANSD	12603	1103
EEVQLAVFGLTANSD	13690	1104
EEVQLLAFGLTANSD	16286	1105
EEVQLLVAGLTANSD	11480	1106
EEVQLLVFALTANSD	18254	1107
EEVQLLVFGATANSD	19978	1108
EEVQLLVFGLAANSD	18863	1109
EEVQLLVFGLTTNSD	20021	1110
EEVQLLVFGLTAASD	19200	1111
EEVQLLVFGLTANAD	17928	1112
EEVQLLVFGLTANSA	22206	1113
EEVQLLVFGLTANSD	18721	1114
THLLQGQSLTLTLES	7756	1115
AHLLQGQSLTLTLES	8602	1116
TALLQGQSLTLTLES	6931	1117
THALQGQSLTLTLES	7683	1118
THLAQGQSLTLTLES	7701	1119
THLLAGQSLTLTLES	4578	1120
THLLQAQSLTLTLES	8471	1121
THLLQGASLTLTLES	4238	1122
THLLQGQALTLTLES	8659	1123
THLLQGQSATLTLES	4430	1124
THLLQGQSLALTLES	8158	1125

THLLQGQSLTATLES	4380	1126
THLLQGQSLTLALES	11699	1127
THLLQGQSLTLTAES	862	1128
THLLQGQSLTLTLAS	2596	1129
THLLQGQSLTLTLEA	5849	1130
THLLQGQSLTLTLES	6545	1131
THLLQGQSLTLTLES	4787	1132
AHLLQGQSLTLTLES	5826	1133
TALLQGQSLTLTLES	5012	1134
THALQGQSLTLTLES	5059	1135
THLAQGQSLTLTLES	5120	1136
THLLAGQSLTLTLES	2956	1137
THLLQAQSLTLTLES	6393	1137
THLLQGASLTLTLES	1933	1139
THLLQGQALTLTLES	5151	1140
THLLQGQSATLTLES	1391	1141
THLLQGQSLALTLES	4749	1142
THLLQGQSLTATLES	813	1143
THLLQGQSLTLALES	8147	1144
THLLQGQSLTLTAES	797	1145
THLLQGQSLTLTLAS	2193	1146
THLLQGQSLTLTLEA	7984	1147
THLLQGQSLTLTLES	5947	1148
Empty (control)	569	

Panel G

PEPTIDE	COUNTS	SEQ ID NO:
GEQVEFSFPLAFTVE	20691	1149
AEQVEFSFPLAFTVE	18546	1150
GAQVEFSFPLAFTVE	17733	1151
GEAVEFSFPLAFTVE	17500	1152
GEQAEFSFPLAFTVE	14764	1153
GEQVAFSFPLAFTVE	16668	1154
GEQVEASFPLAFTVE	6793	1155
GEQVEFAFPLAFTVE	21681	1156
GEQVEFSAPLAFTVE	7767	1157
GEQVEFSFALAFTVE	20480	1158
GEQVEFSFPAAFTVE	10024	1159
GEQVEFSFPLTFTVE	17397	1160
GEQVEFSFPLAATVE	10130	1161
GEQVEFSFPLAFAVE	20627	1162
GEQVEFSFPLAFTAE	18797	1163
GEQVEFSFPLAFTVA	18371	1164
GEQVEFSFPLAFTVE	17662	1165
GEQVEFSFPLAFTVE	19190	1166
AEQVEFSFPLAFTVE	18042	1167

GAQVEFSFPLAFTVE	18079	1168
GEAVEFSFPLAFTVE	19756	1169
GEQAEFSFPLAFTVE	13000	1170
GEQVAFSFPLAFTVE	13930	1171
GEQVEASFPLAFTVE	6533	1172
GEQVEFAFPLAFTVE	20072	1173
GEQVEFSAPLAFTVE	7378	1174
GEQVEFSFALAFTVE	19480	1175
GEQVEFSFPAAFTVE	10589	1176
GEQVEFSFPLTFTVE	18318	1177
GEQVEFSFPLAATVE	9572	1178
GEQVEFSFPLAFAVE	19516	1179
GEQVEFSFPLAFTAE	16765	1180
GEQVEFSFPLAFTVA	18187	1181
GEQVEFSFPLAFTVE	18219	1182
ZZZZZZDTYICEVED	5017	1183
ZZZZZZDTYICEVEZ	5421	1184
ZZZZZZDTYICEVZZ	2166	1185
ZZZZZZDTYICEZZZ	922	1186
ZZZZZZDTYIZZZZZ	564	1187
ZZZZZZZTYICEVED	3031	1188
EEVQLLVFGLTANSD	23357	1189
EEVQLLVFGLTANSZ	15808	1190
EEVQLLVFGLTANZZ	16496	1191
EEVQLLVFGLTAZZZ	14097	1192
EEVQLLVFGLTZZZZ	16473	1193
EEVQLLVFGLZZZZZ	10516	1194
EEVQLLVFGZZZZZZ	10372	1195
EEVQLLVFZZZZZZZ	7333	1196
EEVQLLVZZZZZZZZ	1098	1197
ZEVQLLVFGLTANSD	16716	1198
ZZVQLLVFGLTANSD	5281	1199
ZZZQLLVFGLTANSD	4310	1200
ZZZZLLVFGLTANSD	1026	1201
ZZZZZLVFGLTANSD	664	1202
ZZZZZZVFGLTANSD	779	1203
ZZZZZZZFGLTANSD	760	1204
ZZZZZZZZGLTANSD	657	1205
EEVQLLVFGLTANSD	18040	1206
THLLQGQSLTLTLES	10850	1207
THLLQGQSLTLTLEZ	10269	1208
THLLQGQSLTLTLZZ	4668	1209
THLLQGQSLTLTZZZ	908	1210
THLLQGQSLTLZZZZ	844	1211
THLLQGQSLTZZZZZ	475	1212
THLLQGQSLZZZZZZ	548	1213
THLLQGQSZZZZZZZ	570	1214
THLLQGQZZZZZZZZ	442	1215

ZHLLQGQSLTLTLES	11445	1216
ZZLLQGQSLTLTLES	11631	1217
ZZZLQGQSLTLTLES	7993	1218
ZZZZQGQSLTLTLES	6887	1219
ZZZZZGQSLTLTLES	3305	1220
ZZZZZZQSLTLTLES	4453	1221
ZZZZZZZSLTLTLES	1086	1222
ZZZZZZZZLTLTLES	1201	1223
THLLQGQSLTLTLES	9756	1224
GEQVEFSFPLAFTVE	18856	1225
GEQVEFSFPLAFTVZ	16222	1226
GEQVEFSFPLAFTZZ	12535	1227
GEQVEFSFPLAFZZZ	11384	1228
GEQVEFSFPLAZZZZ	5846	1229
GEQVEFSFPLZZZZZ	4749	1230
GEQVEFSFPZZZZZZ	2208	1231
GEQVEFSFZZZZZZZ	3277	1232
GEQVEFSZZZZZZZZ	742	1233
ZEQVEFSFPLAFTVE	19736	1234
ZZQVEFSFPLAFTVE	18684	1235
ZZZVEFSFPLAFTVE	12892	1236
ZZZZEFSFPLAFTVE	12166	1237
ZZZZZFSFPLAFTVE	2134	1238
ZZZZZZSFPLAFTVE	1454	1239
ZZZZZZZFPLAFTVE	1391	1240
ZZZZZZZZPLAFTVE	1489	1241
GEQVEFSFPLAFTVE	18867	1242
empty (control)	580	

Example 11

This example characterizes CD4 receptor sequences found to have HIV gp120 binding activity in screening tests. Panel A displays information obtained from sequential replacement of amino acid residues by alaninyl residues. In panel A, a (+) signifies a decrease in binding affinity whereas a (>) indicates that replacement of the residue by an alaninyl residue yields an increase in binding affinity. Sequences are shown with amino-terminus at the top and the carboxyl-terminus at the bottom. Right and left sides are from independent assays.

Panel A.

105-113	116-130	131-145	216-229

D	E	T	G
T	E	H	E
++Y++	V	L	Q
+I+	+Q+	L	+V+
C	+L+	+Q+	+E+
+E+	+L+	G	++F++
+V+	+V+	+Q+	S
+E+	+F+	S	++F++
D	G	+L+	P
	+L .	T	++L++
	T	+L++	A
	A	>T>	++F++
	N	+++L+++	T
	S	++E++	V
	D	S	E

Panel B indicates the effect on binding affinity when successive amino acid residues are deleted, either from the amino-terminus (right side-symbols) or the carboxyl-terminus from the bottom (left side-symbol). A (+) signifies a decrease in binding affinity, and the underlined residues indicate which residue was the last residue to be serially deleted.

Panel B.

105-113	116-130	131-145	216-229
D+	E	T	G
T	E+	H	E
Y	V+	L+	Q+
I	Q++	L+	V+
C	L+++	Q++	E+++
+++E	L+++	G++	F+++
++V	V+++	Q+++	S++++
+E	++++F++++	+++S+++	++++F++++
D	++G	+++L	+++P
	+L	+++T	+++L
	T	+++L	++A
	A	++T	++F
	N	++L	+T
	S	+E	+V
	D	S	E

All publications cited herein are hereby incorporated by reference to the same extent as if each publication were individually and specifically indicated to be incorporated by reference and were set forth in its entirety herein.

While this invention has been described with an emphasis upon preferred embodiments, it will be obvious to those of ordinary skill in the art that variations of the preferred embodiments can be used and that it is intended that the invention can be practiced otherwise than as specifically described herein. Accordingly, this invention includes all modifications encompassed within the spirit and scope of the invention as defined by the following claims.